# The structural principles underlying Mo-insertase functionality and protein complex assembly

Von der Fakultät für Lebenswissenschaften

Von der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

D i s s e r t a t i o n

von Ahmed H. Hassan
aus Alexandria

1. Referent: Professor Dr. Ralf-Rainer Mendel
2. Referent: Professor Dr. Dieter Jahn
eingereicht am: 26.03.2021
mündliche Prüfung (Disputation) am: 31.05.2021

Druckjahr 2021

# Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

## Publikationen

Plate J, Sassen WA, **Hassan AH**, Lehne F, Köster RW, Kruse T S-Sulfocysteine Induces Seizure-Like Behaviors in Zebrafish *Front Pharmacol 10, 122 (2019).*

## Posterbeiträge

**Hassan AH**, Krausze J, Meston D, Blankenfeldt W, Mendel RR, Kruse T. “Identification of the interaction between Mo-insertase proteins in *E.coli*”, PROCOMPAS Ph.D. Symposium, Braunschweig (Deutschland), 20. - 21.06.2018.

**Hassan AH**, Krausze J, Ihling C, Sinz A, Mendel RR, Kruse T. “Identification of the interaction between Mo-insertase proteins”, XI. Molybdenum & Tungsten Enzyme Conference 2019, Potsdam (Deutschland), 07. - 12.07.2019

# Acknowledgements

It is surreal that I planned all these years of whom to thank, and when the time came, I am out of words to express my gratitude to everyone who supported me. Firstly, I would like to thank god for his support and for making everything happen at its time. His blessing and support became meaningful with time which made me understand that everything has a reason and a time, regardless of all what we precisely plan.

I would like to thank my professor and mentor, Ralf Mendel. Thank you for all the support you always gave me and for being the wise person I can always count on when I feel troubled. It has been a real joy to learn from you, and I will always remember the history stories and facts in science that you always told us about. Thank you for opening our eyes to academia by letting us all meet with these interesting lecturers and professors to have these intellectual conversations and learn from them. I also would like to thank professor Dieter Jahn for being my second supervisor. Thank you for your support and kindness. I would like to also thank professor Robert Hänsch and his team for the interesting scientific intellect, seminars as well as gatherings.

Very special thanks goes to Dr. Tobias Kruse for supervision and friendship. Thank you for always helping and being too kind with explaining and sharing ideas as well as accepting new ones. I learned a lot from you and your leadership in the lab. Thank you for your steady support and understanding whenever I needed it especially in this work ''Building a protein complex was like building the pyramids''. I cannot be thankful enough.

To the entire lab (past and present) thank you all for keeping the fun alive, your support, and a good working environment. I am incredibly grateful to have had the opportunity to join such an exciting lab full of great minds. I like to thank Dr. Joern Krausze, or as I always said "the man who knows everything", thank you for your extreme patience and kindness. I would not know half as much as what I know about structural biology if not for you. Thomas and Simon, you two were there from the start. Thank you for the good time, scientific discussions and being the PhD companions I would not have dreamed of. Additional thanks to Thomas, thanks for your comments on my thesis, I will have our pictures together hung in the pyramids as we always planned. Special thanks to Jenni "the cloning god" for the friendship, support and company during very long lab days – Your death metal music kept me awake in late lab evenings. I also like to thank Dr. Dagmar Zwerschke, thank you for all the help whenever I needed it, your kindness and constant support. Big thanks go to Merle, Marius, Linus, Sven and Carla for the crazy and fun times in the lab, and I wish you all are doing well in your new research careers. Distinctive thanks to Maria for always being there and always asking, thank you for the company and the constant care ''You are the best''. Importantly, I thank my students Tamim and Milena.

Working with you guys was one of the best experiences I ever had. I thank you both for keeping up with me and my working ways. I thank Tamim for the coffee discussions "Zekryat" as well as the amazing times in the lab with the excellent scientific discussions. I thank Milena for being part of this work, the fun times in the lab, the music (keeping up with my music taste) and definitely being the best bachelor's student in the world. There are far too many people in the Mendel lab to thank individually. Nevertheless, I thank you all for the enjoyable times, passion for science and collaborative atmosphere.

Beyond Braunschweig, I like to thank professor Andrea Sinz. Thank you for letting me join your lab for some time and do all the crosslinking experiments, teaching me to think analytically and think in "Daltons, not kilo Daltons". I also like to thank Dr. Christian Ihling, your help and patience in teaching me to work with the MS and for your help with data interpretation. I would like to thank professor Petra Wendler for her guidance in letting me learn from her fantastic lab as well as Thomas Beck, who guided me through working on the EM. I like to thank Dr. Mark Carlile, thank you for believing in me since my masters and pushing me to go further in science and yes, I will have my PhD graduation ceremony on "a camel".

Outside science, I thank Amr, Bialy, Omar, Moataz, Shaarawy and the F73 squad. Thank you guys for the continuous support, virtual company and fun. Thank you for supporting me when I thought that doubling up on graduate degrees is undoubtedly a good idea. Exceptional thanks to Daniel, thank you for being the crazy little brother I never had. I would not have been here if it was not for you.

Finally, I thank my family. Mom, for making everything worth it, for the constant teachings, for the sacrifice, for never allowing me to underestimate myself, support, encouragement, and always giving all without taking anything in return. I am forever thankful for you and anything I write here is not enough to explain how grateful I am to have you. Hagar, without such an amazing sister's help and motivation, I might have remained working in a pharmacy still. Our discussions and jokes always made my day. Thank you for keeping up with my craziness and calming me when I am stressed. I will always be grateful for you ''Ahmodetheky one thing''. Elisabeth, thank you for being a reason for me to smile every day and for your constant support "Baby Eli said that Ahi can do it!". Thank you for your patience and your love.

I would like to thank all who doubted me from the beginning since my bachelor's, and these doubts made me who I am now. I would like to express my apologies for those whom I forgot, even though the list was already way too long.

# Summary

Molybdenum (Mo) is a crucial element for the survival of most species, which is due to its vital duty for Mo-dependent enzymes. However, in the cell, Mo is biologically inactive unless it is complexed within a distinctive scaffold referred to as molybdopterin (MPT). Insertion of Mo into MPT yields the physiologically active molybdenum cofactor (Moco) found in the active site of Mo-enzymes. Molybdenum insertases (Mo-insertases) catalyze the Mo-insertion reaction into MPT. These consist of two functional domains referred to as G- and E- domain. With one known exception, in eukaryotes, the two domains are fused together, however, in prokaryotes, they are separate entities. In Escherichia coli, these are annotated as MogA and MoeA, while in *Arabidopsis thaliana* and *Homo sapiens,* the Mo-insertase is named Cnx1 and gephyrin, respectively. Synthesis of Moco relies on the directed interaction of the Mo-insertase E- and G-domain within the Mo-insertase complex, thus ensuring the protein protected metabolite transfer here. However, any structural and biochemical information about the Mo-insertase complex is missing so far. This work presents the first structural insight into the assembly of prokaryotic and eukaryotic E- and G-domains within the Mo-insertase complex. For the MoeA-MogA complex, native mass spectrometry (native MS) gave first insights into the stoichiometry and size of the complex. Crosslinking based mass spectrometry (XL-MS) and computational modelling subsequently revealed how MogA interacts with MoeA, which was confirmed by negative staining electron microscopy (EM). For the eukaryotic Mo-insertases Cnx1 and gephyrin, a novel purification approach was established, ensuring both recombinant proteins to be co-purified with minimal degradation products. My XL-MS based structural proteomics approach revealed the Cnx1 G-domain to bind to two distinctive interaction sites on the E-domain, which ultimately results in the formation of a closed complex consisting of 3 E-domain dimers and 2 G-domain trimers. Other stoichiometries may occur, assuming that in cellula, G-domain trimers exist as separate entities. Formation of these may result from Cnx1 degradation, which allows it to speculate that in cellula Mo-insertase complex formation underlies dynamic and potentially demand driven processes. Further insights into the Cnx1 structure was obtained by an combined structural proteomics ab initio modelling approach of the linker which revealed consistency with respect to the Cnx1 closed complex model presented here. In order to further substantiate the obtained results for the plant Mo-insertase Cnx1, structure-guided mutagenesis was done, targeting potentially critical residues within the interaction interfaces determined. As an important finding, recombinant biochemistry revealed variants with an impaired metabolite transfer capacity. Notably, any of these variants was traced back to one of the two identified interaction sites, which thus was confirmed as a “metabolic interaction” site and accordingly, the second interaction site was identified as a “structural interaction” site. Consistent results were obtained for the human gephyrin, as negative staining EM also reveals two defined G-domain interaction sites with the E-domain.

# Table of contents

# List of Figures

# List of Tables

# Abbreviations

| Acroynm | Definition |
|---|---|
| 3′,8-cH2GTP | 3′,8-cyclo-7,8-dihydro-guanosine 5′-triphosphate |
| aa | amino acid |
| ADP | Adenosine diphosphate |
| Amp | ampicillin |
| AO | Aldehyd oxidase |
| AP | alkaline phosphatase |
| ATP | Adenosine triphosphate |
| BiFC | Bimolecular fluorescence complementation |
| cPMP | Cyclic pyranopterin monophosphate |
| Cryo-EM | Cryogenic electron microscope |
| Da | Dalton |
| DFDNB | 1,5-difluoro-2,4-dinitrobenzene |
| dm | double mutant |
| DMF | Dimethylformamide |
| DMSO | Dimethylsulfoxide |
| DSBU | Disuccinimidyl dibutyric urea |
| DSS | Disuccinimidyl suberate |
| DSS | Disuccinimidyl suberate |
| DSSO | Disuccinimidyl sulfoxide |
| DST | Disuccinimidyl tartrate |
| DTT | dithiothreitol |
| E. coli | *Escherichia coli* |
| EDC | 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride |
| EDTA | ethylenediaminetetraacetic acid |
| Em/em | Electron microscope |
| FA | Formic acid |
| $GABA_A$ | γ-aminobutyric acid type A |
| GTP | Guanosine triphosphate |
| H. sapiens | *Homo sapiens* |
| HADDOCK | high ambiguity driven protein-protein docking |
| HCl | Hydrogen chloride |
| hr | Hour/hours |
| HPLC | high-performance liquid chromatography |
| IAA | iodoacetamide |
| IEX | Ion exchange chromatography |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| IPTG | Isopropyl-β-D-thiogalactopyranoside |
| kDa | Kilo Dalton |
| KLD | kinase-ligase-DpnI |
| LB media | Lysogeny broth media |
| LB-Amp | LB-media containing ampicillin |

| | |
|---|---|
| MCD | MPT cytosine dinucleotide |
| MCP | Moco carrier protein |
| min | minute/minutes |
| $MgCl_2$ | Magnesium chloride |
| MGD | MPT guanosine dinucleotide |
| $MgSO_4$ | Magnesium sulfate |
| Midiprep | large-scale plasmid DNA preparation |
| Miniprep | small-scale plasmid DNA preparation |
| Mo | Molybdenum |
| MoBP | Moco binding protein |
| MoCD | Moco – deficiency |
| Moco | Molybdenum cofactor |
| Mo-insertase | Molybdenum insertase |
| MPT | Molybdopterin |
| MPT-AMP | Adenylated MPT |
| MS | Mass spectrometry |
| $Na_2MoO_4$/ NaMo | sodium molybdate |
| Native MS | Native mass spectrometry |
| NHS | N-hydroxysuccinimide |
| NMR | Nuclear magnetic resonance |
| NR | Nitrate reductase |
| PCR | polymerase chain reaction |
| PDE1 | phosphodiesterase 1 |
| PVDF | polyvinylidene fluoride |
| RDA | Recommended Dietary Allowance |
| RMSD | Root-mean-square deviation |
| SAM | S-adenosylmethionine |
| SAXs | Small-angle x-ray scattering |
| SDS-PAGE | sodium dodecyl sulfate-polyacrylamide |
| SEC | Size exclusion chromatography |
| sec | Seconds |
| SO | Sulfite oxidase |
| SOC | Super Optimal broth with Catabolite repression |
| SSC | S-sulfocystein |
| Strep | Streptactin |
| TAE | Tris-Acetate-EDTA |
| TBS-T | Tris-buffered saline-Tween |
| TFA | trifluoroacetic acid |
| A. thaliana | *Arabidopsis thaliana* |
| Tris | Tris-(hydroxymethyl)-aminomethane |
| v/v | Volume per volume |
| w/v | Weight per volume |
| w/w | Weight per weight |
| XDH | Xanthine dehydrogenase |
| XL-MS | Crosslinking Mass Spectrometry |

# Chapter One

# 1. Introduction

## 1.1. Molybdenum and Molybdenum Cofactor

### 1.1.1. Overview

In nature, Molybdenum (Mo) is one of the 90 naturally existing elements and it is classified as a transition element, or a metal, in the periodic table (Mendel, 2013). Mo is also known to be a ''refractory metal'' of extraordinary heat and wear resistance. Thus it is widely used in high-temperature furnaces, material testing equipment, aerospace gears and defense components (Lunk *et al.*, 2017). Mo can be ubiquitously found everywhere on earth existing as $MoO_4$ or $MoS_2$, the crustal abundance of Mo is equivalent to 1-15 ppm, while 10 µg/L in sea water, as the most widespread transition metal in open sea water (Smedley *et al.,* 2017). Most species rely on Mo for normal biological functions, including energy production and waste products removal. Nevertheless, Mo is needed in minuscule amounts for the various organisms (Mendel, 2013). The first hint on the distribution of Mo in different species goes back to 1932, by the work of Meulen, in which Mo was acknowledged as an essential metal for various organisms (Meulen, 1932). For humans, 45 ug/day is the daily Recommended Dietary Allowance (RDA), which is easily obtained in the everyday diet (FDA, 2016). Although it is doubtful to have excessive Mo intake above 2 mg/day, Mo is lethal above the tolerable levels for different species (ATSDR, 2020). Notably, Mo is not biologically active or readily used by different organisms. Conversely, it needs to be complexed in a unique scaffold, referred to as metal-containing petrin (MPT), which forms the biologically active molybdenum cofactor (Moco) (Mendel, 2013; Schwarz *et al.*, 2009; Mendel & Kruse, 2012).

## 1.1.2. A brief history of the molybdenum cofactor

In 1963, the first experimental evidence on the existence of Moco was observed while characterizing gene mutations in *Aspergillus nidulans* by Cove and Patman (Cove & Patman, 1963). Remarkably, in the late 60s Mudd, showed for the first time the impact of sulfite oxidase deficiency in humans (Mudd *et al.*, 1967), while in 1971, sulfite oxidase was characterized as a Mo- containing protein (Cohen *et al.*, 1971). In 1976, the first Moco-mutant plants were generated (Mendel & Müller, 1976), while the first Mo-repair of cnx mutants was shown in 1981 by R. R. Mendel (Mendel *et al.,* 1981). The eukaryotic Moco biosynthesis prevailed from 1985 to 1992 (Mendel *et al.,* 1986; Müller & Mendel, 1989) (Figure 1). From there on, numerous studies were published that deciphered the ambiguities of Moco and Moco biosynthesis. These studies included the characterization of Moco intermediates, Moco-dependent enzymes, and Moco biosynthesis proteins as well as their structures (Mendel, 2013; Schwarz *et al.*, 2009; Mendel & Kruse, 2012). Up till now, although we have a good understanding about the relevancy of Mo and Moco in different organisms, however, there are still plenty of mysteries regarding Moco and Mo in biological systems.

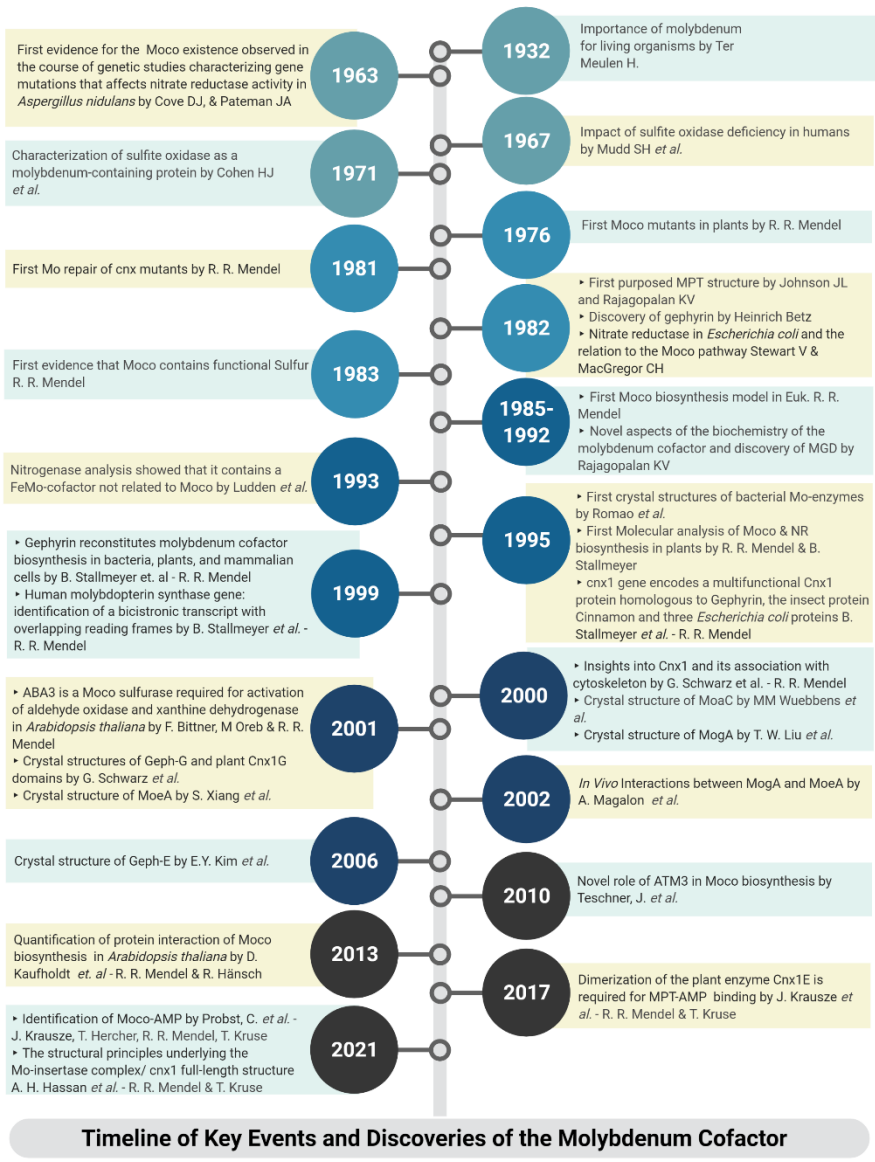



**Figure 1. Timeline of key events and discoveries of Moco.** Timeline showing fundamental discoveries and events which are the foundation of our current knowledge about Moco.

## 1.1.3. The significance of the molybdenum cofactor

The importance of Moco is due to the inevitable involvement of Mo in all Mo-dependent enzymes. These enzymes are essential for different biochemical processes in the overall metabolism of carbon, nitrogen and sulfur (Mayr *et al.*, 2020). With the exclusion of nitrogenase, different variants of Moco form part of the active center of Mo-dependent enzymes needed for the different species. There are four Mo-dependent enzymes out of

seven known in eukaryotes are found in humans, while five enzymes are present in plants (Figure 2). These enzymes are sulfite oxidase (SO), mitochondria amidoxime-reducing component (mARC), xanthine oxidase (XO) and Aldehyde oxidase (AO) (Mendel, 2013).

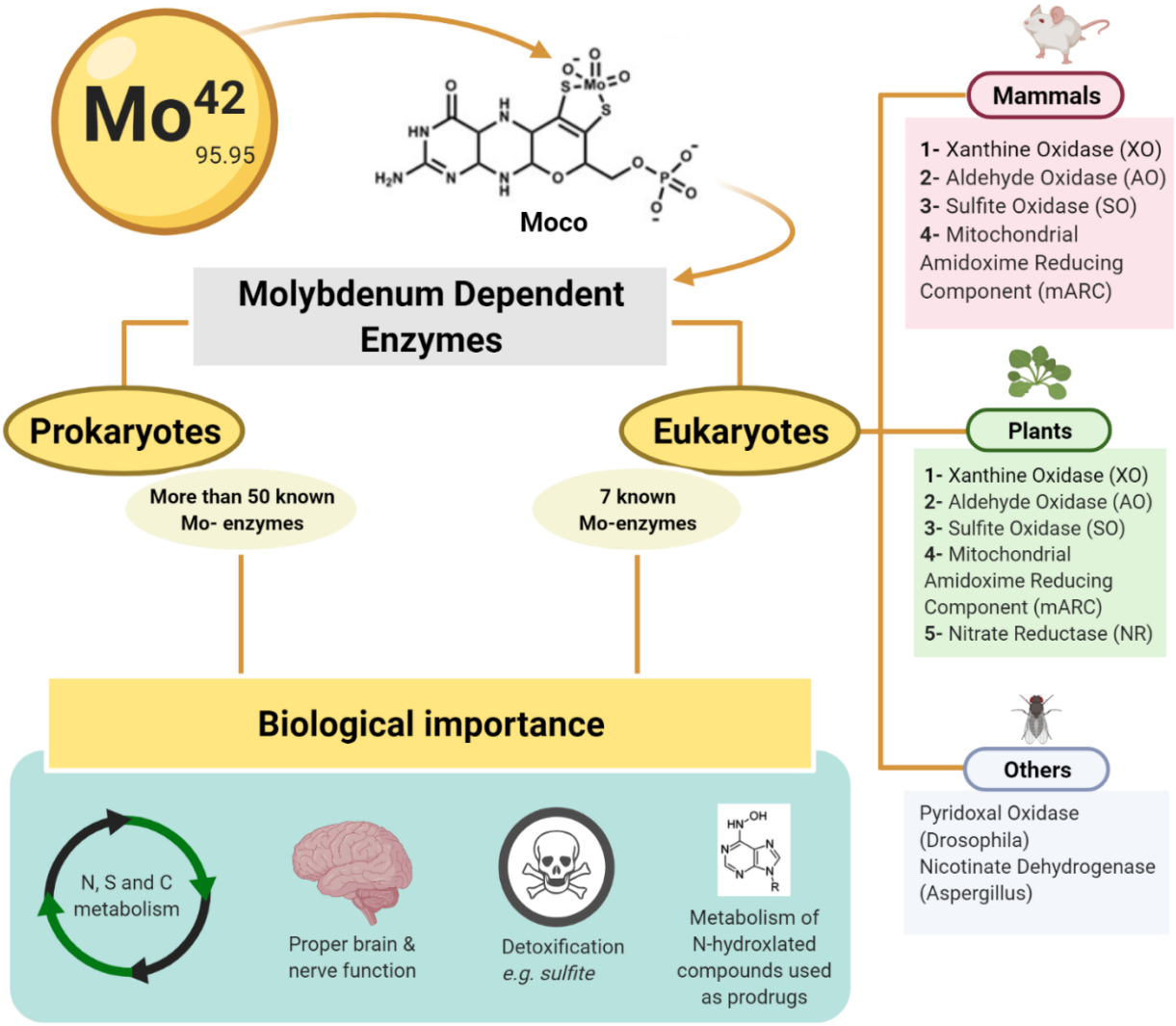



**Figure 2. Moco dependent enzymes and their biological importance.** The transition metal Mo with atomic number 42 is not biologically active unless it is complexed in a scaffold that forms Moco. In prokaryotes, there are more than 50 known Mo-dependent enzymes. However, in eukaryotes, seven enzymes are known. Sulfite oxidase (SO), mitochondria amidoxime-reducing component (mARC), xanthine oxidase (XO) and Aldehyde oxidase (AO) are the four Mo-depended enzymes needed for humans (red). For plants, XO, AO, SO, mARC and nitrate reductase (NR) are the five Mo-dependent enzymes needed. Additionally, pyridoxal oxidase in drosophila melanogaster and nicotinate dehydrogenase in Aspergillus nidulans. Mo-dependent enzymes are essential for the different species for their nitrogen, sulfur and carbon as well as the detoxification of many metabolites such as sulfite by sulfite oxidase. Moreover, proper brain and nerve function and metabolism of N-hydroxylated compounds which are used as prodrugs require Moco-dependent enzyme function

While the two isoforms of mARC are important in the metabolism of N-hydroxlated compounds used as prodrugs, SO, which localize in the intermembrane space of the mitochondria, is involved in the crucial oxidation of sulfite into sulfate (Havemeyer *et al.*, 2011; Hille, 1996). Moreover, XO is crucial for purine catabolism by converting hypoxanthine into xanthine, which is subsequently converted into uric acid (Keyelle et al. 2010; Hille, 1996). The oxidation of different heterocyclic aldehydes into carboxylic acids to release hydrogen peroxide and superoxide anions rely on AO (Terao *et al.*, 2020). In addition to the four enzymes required by humans, nitrate reductase (NR) is needed for plants, algae and fungi. NR is critical for these plants as it reduces nitrate to nitrite, which is the most important source of nitrogen in the soil utilized by NR in different mechanisms (Wajmann *et al.,* 2020; Mendel, 2013).

Consequently, any gene defect in the Moco biosynthesis would result in Moco – deficiency (MoCD) which is fatal for most organisms because of the improper functionality of these enzymes (Schwarz, 2016). For instance, in humans, malfunctional Moco biosynthesis leads to severe neurological symptoms, seizures, dislocation in the ocular lenses as well as other symptoms, which ultimately results in the death of the affected patients (J. Reiss *et al.,* 2011). There are three known types of MoCD known, deficiency of MOCS1 as type A, deficiency of MOCS2 or MOCS3 as type B and deficiency of gephyrin as type C, which all result in improper function of Mo-dependent enzymes (Bayram *et al.*, 2013; Schwarz, 2016). Most importantly, the defect in SO would promote the accumulation of toxic sulfite. The underlying molecular reason for the MoCD phenotype is S-sulfocystein (SSC) formation, a glutamate orthologue that is neurotoxic, which causes mental retardation in humans and seizure-like behaviour effects in zebrafish (Plate *et al.* 2019).

## 1.1.4. Molybdenum cofactor biosynthesis

The biosynthesis of Moco is conserved, which could be tracked to the most primaeval biological processes known to life (Zhang and Gladyshey, 2008). Even though Moco biosynthesis is highly conserved among bacteria, plants and humans, however, they are relatively diverse on the gene level (Mayr *et al.,* 2020). In bacteria and archaea, eight genes are required for the biosynthesis of Moco. On the other hand, six genes are required in plants and only four in humans. On the protein level, both plants and humans have a total of six

proteins regardless of the number of genes encoding them (Mendel, 2013) (Figure 3). Another two proteins are known to be involved in the Moco biosynthesis in bacteria and archaea for the additional step in order to generate MPT cytosine dinucleotide (MCD) or MPT guanosine dinucleotide (generates bis-MGD) (Leimkuhler, 2020).

In eukaryotes, the three-step biosynthesis of Moco starts in the mitochondria with the aid of two proteins, MOCS1A and MOCS1B in humans, Cnx2 and Cnx3 in plants, while MoaA and MoaC in bacteria. The two proteins help in the conversion of guanosine triphosphate (GTP) into cyclic pyranopetrin monophosphate (cPMP), which is then exported outside of the mitochondria and converted to MPT (Mendel & Kruse, 2012). MoaA is part of the superfamily of radical S-adenosylmethionine (SAM), and it is characterized by [4Fe-4S] cluster that cleaves SAM producing adenosyl radical, which forms a +2 oxidized form of the [4Fe-4S] (Yokoyama *et al.,* 2018). SAM in both MoaA and MOCS1A is solvent accessible and not profoundly buried within the active site of the protein. Although MoaA was known to play the central role in the pyranopterin formation and for the synthesis of the intermediate 3′,8-cyclo-7,8-dihydro-guanosine 5′-triphosphate (3′,8-cH2GTP), structural and functional characterization of MoaC showed the crucial involvement of MoaC in the conversion of 3′,8-cH2GTP to cPMP (Yokoyama, 2018).

The conversation of cPMP to MPT is done by the MPT-synthesis complex in a two subsequent sulfur transfer reaction. The MPT-synthesis is composed of two small subunits of MOCS2A and two large subunits of MOCS2B in humans. In plants, MOCS2A and MOCS2B are referred to as Cnx6 and Cnx7, while MoaD and MoaE in bacteria, respectively (Rudolph *et al.,* 2001; Mendel, 2013). Initially, cPMP binds to MoaE (Cnx6 or MOCS2B) to establish the first thiol, which is near the MoaD (Cnx7 or MOCS2A) double glyine motifs. This leads to ring-opening of the cyclic phosphate, then introducing the first thiol and subsequently, the second thiol by MoaD to form the dithiolene moiety. The resulferation of MoaD is facilitated by the sulfarase protein MOCS3, Cnx5 in plant or MoeB in bacteria, in which a sulfur atom is provided in an ATP and $Mg^{+2}$ dependent manner (Rudolph *et al.,* 2001; Wuebbens *et al.,* 2003).

Finally, a two-domain protein in eukaryotes or the separately expressed MoeA and MogA in prokaryotes (Cnx1 in plants and gephyrin in humans) are responsible for adenylating MPT, in an ATP-$Mg^{+2}$ dependent manner (Krausze *et al.,* 2017). The result is the formation of MPT-

AMP (Kuper *et al.,* 2004). Subsequently, the removal of the AMP and insertion of molybdate, which results in the formation of Moco. Afterwards, Moco is either 1) sulfurated by ABA3 in plants and HMCS in humans for the transfer to XO or AO (Ichida *et al.,* 2001; Heidenreich *et al.,* 2005), or 2) carried by a Moco carrier protein (MCP) or a Moco binding protein (MoBP) (Kruse *et al.,* 2010; Krausze *et al.,* 2020), or 3) possibly handed over directly to the Moco user enzymes. While Moco biosynthesis is prevalent amongst all species, surprisingly, some few eukaryotic species are known to lack Moco biosynthesis completely. Prominent examples are *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (Zhang & Gladyshev, 2008).

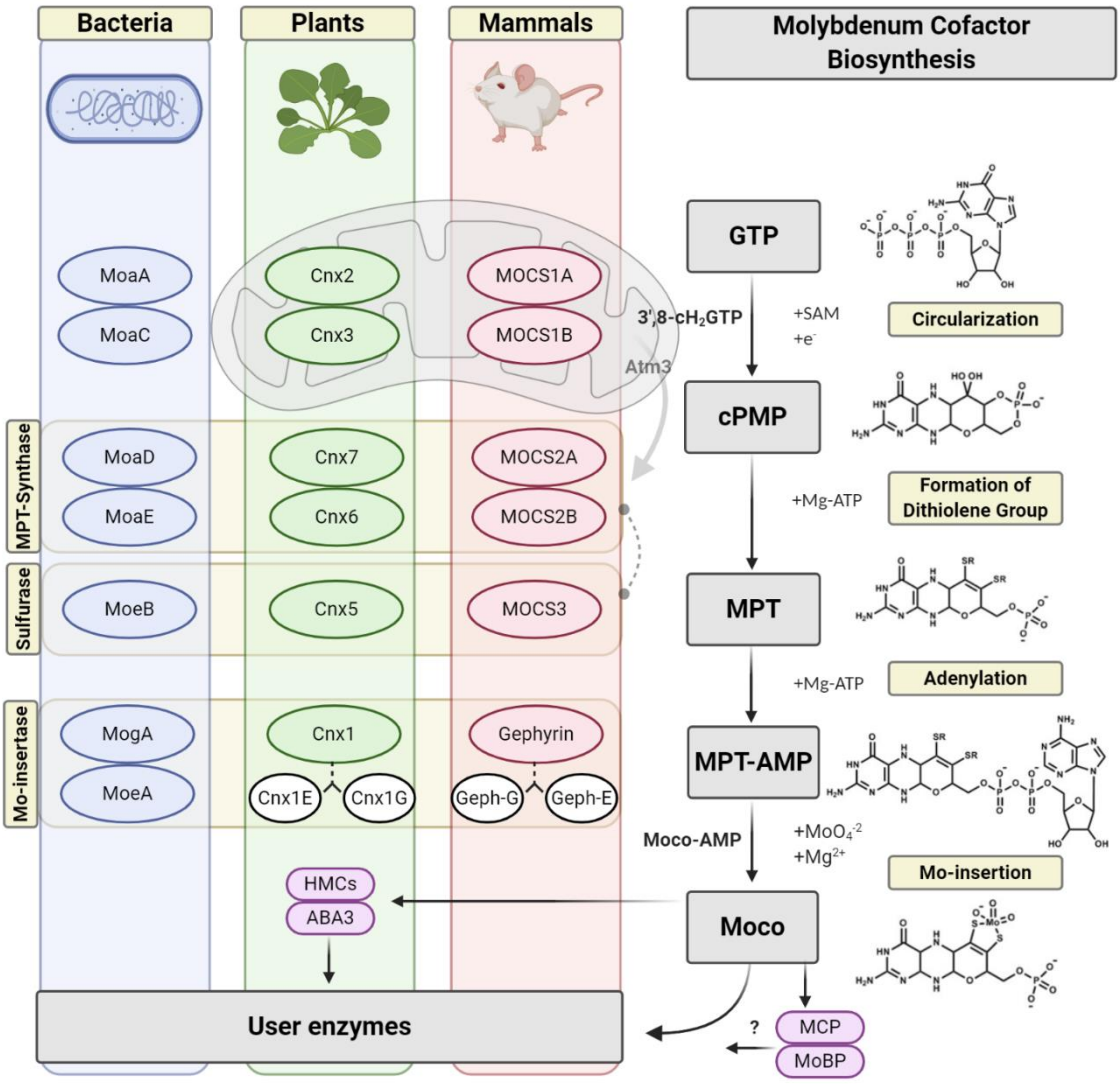



**Figure 3. The Moco biosynthesis.** The biosynthetic pathway is divided into three steps which start in the mitochondria, shown on the right (schematic gray mitochondrium). On the left, the names of the proteins

in humans (red), plants (green) and bacteria (blue) involving in the respective step. In eukaryotes, the first step is the circularization of GTP to cPMP in the mitochondria, which includes an intermediate 3',8-cH$_2$GTP product in a Mg-ATP dependent manner. In the second step, after export to the cytoplasm via ATM3 (plants), cPMP is converted to MPT by the aid of heterotetrameric MPT-synthase complex and the dithiolene group is formed. The MPT sulfurase is responsible for reloading the small subunits of the MPT-synthase complex with sulfur. In the third step, MPT is adenylated to form MPT-AMP by the G-domain of the two-domain Mo-insertase protein. Subsequently, MPT-AMP is deadenylated and Mo is inserted by the E-domain of the Mo-insertase. Moco-AMP is a possible intermediate form in the last step of biosynthesis. Once Moco is formed, it can be either directly transferred to the user enzymes or sulfurated by HMCs in humans & ABA3 in plants (purple) or alternatively indirectly transferred by a MCP (algea and some cyanobacteria) or MoBP (plants).

## 1.2. Mo-insertases and the final step of Molybdenum cofactor biosynthesis

### 1.2.1. Overview

The final step in the eukaryotic Moco biosynthesis is catalyzed by a multi-domain protein that combines the function of two separate bacterial proteins. This protein is referred to as Mo-insertase, due to the fact that the aid of this protein inserts Mo. As evolution took its course, two different gene fusion events occurred to combine both G- and E- domains of the Mo-insertases (Mendel, 2013; Krausze *et al.,* 2017). In general, the crystal structures obtained of G-domains from different species showed the formation of trimers, while dimer formation for the E-domain in solution. The overall structure of the different G- and E-domains is highly similar (Figure 4), with slight differences in the individual monomers of different species (Schwarz *et al.,* 2001; Kuper *et al.,* 2004; Kasaragod & Schindelin, 2016). Similarly, on the sequence level, Mo-insertases tend to have high sequence similarity and plenty of strictly conserved residues (Krausze *et al.,* 2018).

Cnx1 in plants is a fusion of both domains via a linker with N-terminal E-domain and C-terminal G-domain orientation. On the other hand, gephyrin in humans has a longer linker and the reverse orientation of domains, N-terminal G-domain and C-terminal E-domain (Schwarz *et al.,* 2001). The initial Moco biosynthesis steps were extensively studied and explained, nonetheless, the final step is enigmatic, specifically the mechanism behind molybdate insertion into the MPT-AMP dithiolene. The final step of Moco biosynthesis is divided into two reaction steps. Initially, the adenylation of MPT and subsequently, Mo-insertion and MPT-AMP hydrolysis to release Moco (Hercher *et al.,* 2020). In the first stage, the MPT, bound to a highly conserved binding site in the G-domain, is adenylated with the aid of $Mg^{2+}$ and ATP. The binding site consists of three parts, a proline-glycin motif, two

aspartates and a glycine-glycine-threonine motif, remarkably, all three parts are functionally important (Llamas *et al.,* 2006; Kuper *et al.,* 2004; Kuper *et al.,* 2003).
The subsequent reaction step involves the E-domain, and it includes two sub-steps. At first, the insertion of Mo in the MPT-AMP dithiolene moiety and then the hydrolysis of the MPT-AMP, which results in the cleavage of the AMP (Krausze *et al.,* 2017; Hercher *et al.,* 2020). Furthermore, recent works prevailed the presence of a potential novel intermediate, which is referred to as Moco-AMP (Probst *et al.,* 2021). Thus formed Moco is physiologically active.

In contrast to eukaryotes, prokaryotes have an additional step, including the attachment of a nucleotide moiety to Moco to form MGD. This last additional step is done with the aid of two co-transcribed proteins, namely MobA and MobB (Mendel, 2013; Leimkuhler, 2020). In bacteria such as *Escherichia coli*, there are different Moco forms for the various enzyme families, including a di-oxo Moco for sulfite oxidase family, sulfurated molybdopterin cytosine dinucleotide cofactor (MCD) for the xanthine oxidase family, and finally a bis-MGD for the DMSO reductase family (Leimkuhler, 2020; Iobbi-Nivol & Leimkühler, 2013).

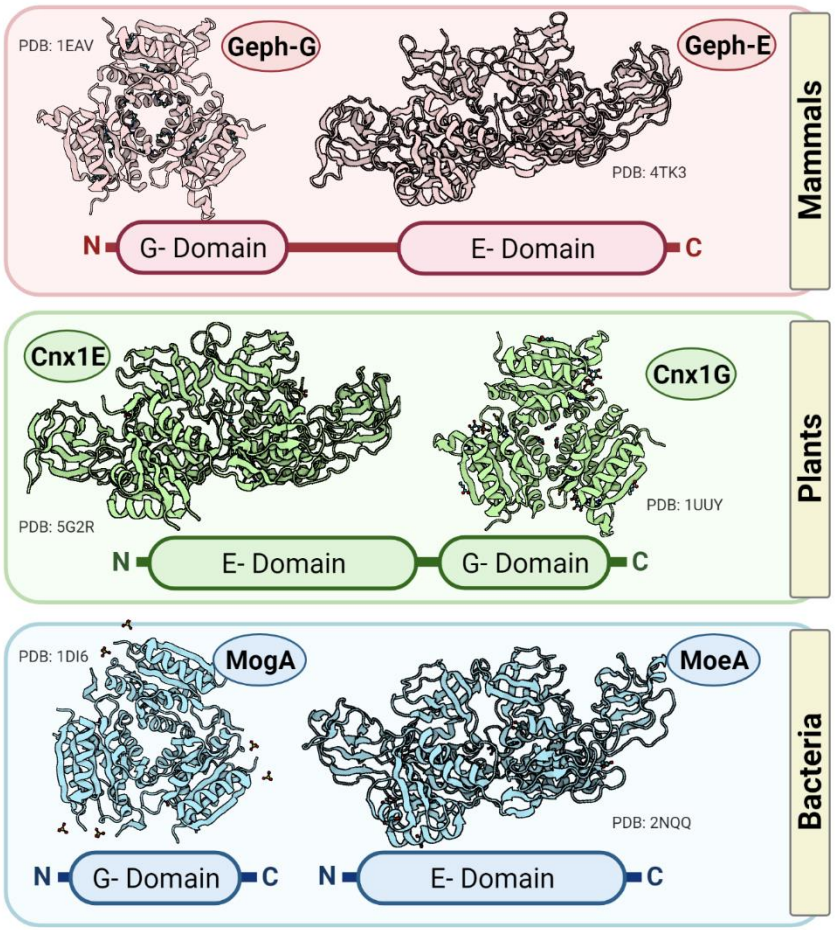



**Figure 4. Mo-insertases in mammals, plants and bacteria.** Throughout the different known species, the Mo-insertases are composed of a homotrimeric G-domain and a homodimeric E-domain. The structures are very similar when comparing bacteria, plants and mammals. In bacteria (blue) the G-domain MogA (PDB: 1DI6) and E-domain MoeA (PDB: 2NQQ) are separately expressed. In plants, the fused N-terminus E-domain ''Cnx1E'' (PDB: 5G2R) and the C-terminus G-domain ''Cnx1G'' (PDB: 1UUY) form Cnx1. Gephyrin in mammals are composed of N-terminus G-domain ''Geph-G (PDB: 1EAV) and C-terminus E-domain ''Geph-E'' (PDB: 4TK3).

## 1.2.2. MogA and MoeA

In prokaryotes, both MogA and MoeA are responsible for the Moco formation. The two proteins are known to have a function solely related to Moco biosynthesis, unlike eukaryotes (Nichols & Rajagopalan, 2002). The inevitable interaction between MogA and MoeA was first described using a bacterial two-hybrid assay. This study also demonstrated an interaction between the MPT-synthase complex, MoaD and MoaE complex with the Mo-insertase complex in the Moco biosynthesis of bacteria (Magalon *et al.*, 2002). Despite the fact that MogA and MoeA have been proven to bind MPT with different affinities, MPT handover by the MPT-synthases complex directly to MogA would be expected (Schwarz *et al.,* 1997; Kuper *et al.,* 2000). Subsequently, MPT is adenylated, handed over to MoeA and

Mo is inserted to release Moco. Additionally, MoeA and MogA are known to interact with MobB and MobA from the subsequent step to form bis-MGD used by DMSO reductase family enzymes (Magalon *et al.*, 2002) (Figure 5).

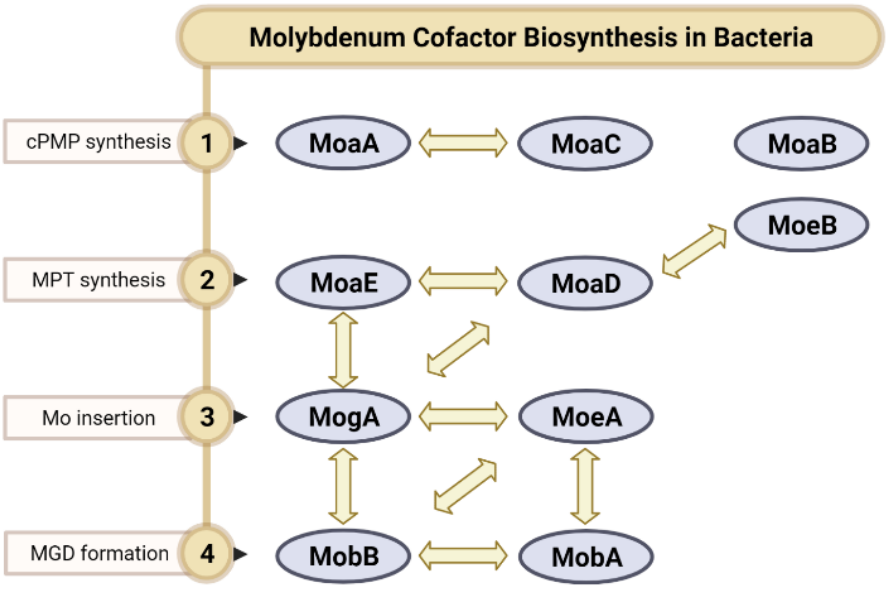



**Figure 5. Moco biosynthesis in bacteria and protein-protein interaction in the pathway.** The biosynthetic pathway in bacteria is composed of 4 steps. Nomenclatures of enzymes is described according to *E. coli.* In each of the steps, a direct protein-protein interaction is shown with an arrow (yellow). MoaA interacts with MoaC from the first step. In the second step, MoaE interacts with MoaD, which interacts with MoeB. Both MoaE and MoaD interacts with MogA from the third step. All proteins from the last two steps were shown to interact with each other.

## 1.2.3.Cnx1

Similarly to bacteria, in plants, Cnx1 resemble a part of Moco biosynthetic complex including other three cytosolic proteins namely Cnx5, Cnx6 and Cnx7 in order to protect the delicate intermediates (Kaufholdt *et al.* 2016). As shown in (figure 6), Cnx2 and Cnx3 are interaction partners responsible for the first step of the biosynthesis. These insights were shown using *in vivo* interaction methods including bimolecular fluorescence complementation (BiFC) as well as split-luciferase assay (Kaufholdt *et al.* 2016). Furthermore, in the same way that its mammalian homolog gephyrin does in neuronal cells, Cnx1 binds

to the cytoskeleton, signifying a hypothetical model for the Moco-biosynthetic anchoring mechanism by Cnx1 in plant cells (Stallmeyer *et al.,* 1999; Schwarz *et al.,* 2000).

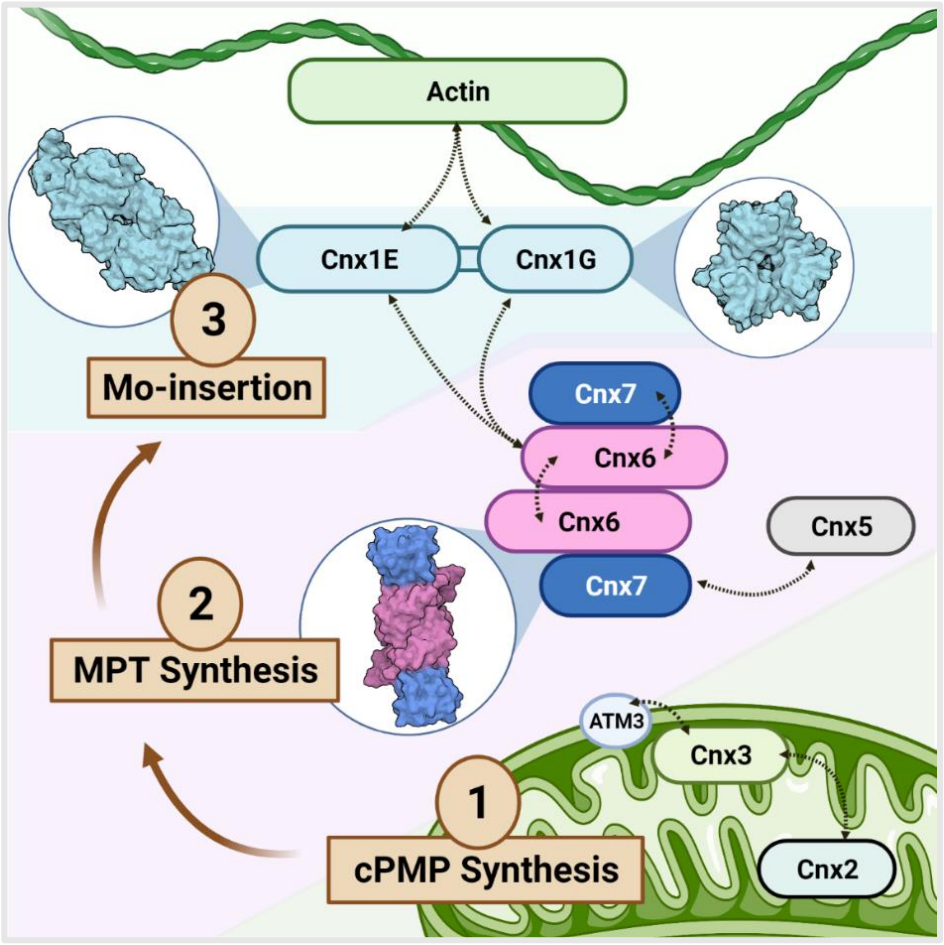



**Figure 6. Moco biosynthesis in plants, including the protein interactions involved.** Dashed arrows indicate direct protein-protein interaction within the Moco biosynthesis complex in plants (Interactions strength is variable). The biosynthesis starts in the mitochondria with cPMP synthesis by Cnx2 and Cnx3. Next, MPT synthesis requires Cnx6 (pink) and Cnx7 (dark blue). Cnx7 interacts with the sulfurase Cnx5 (grey). The final step of Mo-insertion is catalyzed by the two-domain protein Cnx1 which is composed of the trimeric Cnx1G (cyan) and the dimeric Cnx1E (cyan). Both domains interact with Cnx6 and actin (green).

## 1.2.4.Gephyrin

In 1982, gephyrin was discovered by Heinrich Betz's group as an unknown protein co-purified with glycine receptors (Pfeiffer *et al.,* 1982). From then on, gephyrin was extensively studied from the cellular and biochemical perspective. Astonishingly, in the 90s, gephyrin was identified as the homolog of Cnx1, the insect protein ''cinnamon'' and three other *Escherichia coli* proteins (Kamdar *et al.,* 1994). Henceforth, the biochemical analysis showed that gephyrin could reconstitute Moco biosynthesis in bacteria, plants, and mammalian cells. These studies showed that gephyrin is the Mo-insertase in mammals and that it has significance in neural functions and particular importance in Moco biosynthesis (Stallmeyer *et al.,* 1999).

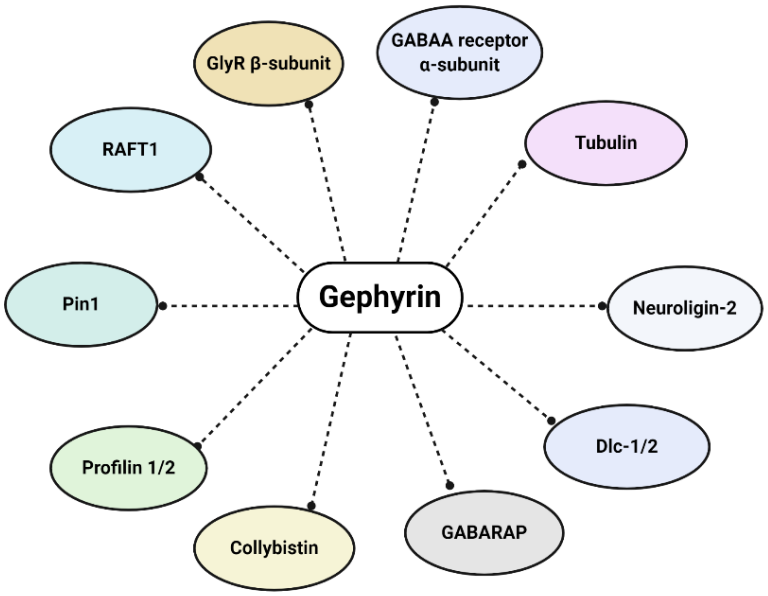



**Figure 7. Gephyrin protein- protein interaction partners in the nervous system.**

For organizing the inhibitory postsynaptic network, gephyrin is the most comprehensively studied scaffold, which is crucial for the clustering of γ-aminobutyric acid type A ($GABA_A$) and glycine receptors. The function of gephyrin in clustering receptors is of absolute importance for the transmission of the inhibitory synapse (Tyagarajan & Fritschy, 2014). In addition, gephyrin was shown to bind with numerous classes of inhibitory synaptic entities (Figure 7) (Choii & Ko, 2015).

Interestingly, most of the interaction partners are interacting directly with the linker region or the E-domain. The different protein-protein interactions with other partners proved the essential role of gephyrin for normal synaptic function, plasticity, and formation (Tyagarajan & Fritschy, 2014). In cultured rat neurons, gephyrin accumulation was observed at the developing postsynaptic membrane specialization, which showed aggregates that were visualized by immunostaining (Kirsch *et al.,* 1993). Interactions of gephyrin with

cytoskeleton elements as well as glycine receptor subunit β were proven to be of absolute importance for the localization of glycine receptor proteins (Allison *et al.,* 2000). Consequently, gephyrin knock-out mice die soon after birth with symptoms resulting from non-functional glycine receptors (Kneussel *et al.,* 1999). Although advancement has been made with our understanding of gephyrin and its roles yet, the protein remains a mystery in many different aspects, primarily due to the linker. The linker region in gephyrin is known to be necessary for a normal receptor anchoring function. Most of the identified interactions with gephyrin include the linker region and/or the E-domain.

Due to the high flexibility of the long linker region, there is no structure of the linker or the full-length gephyrin or cnx1. Additionally, post-translational modifications (PTMs), which probably induce conformational changes that influences the structure of the linker as well as the G- and E- domains (Fritschy *et al.,* 2008). There are 22 phosphorylation sites known in the linker region and one site in the E-domain as well as palmitoylation and other modifications (Choii & Ko, 2015). Although there is no crystal structure known for the full-length gephyrin and the complex, anongoing hypothesis of gephyrin clustering in the neurons was assumed (Kneussel & Betz*,* 2000; Groeneweg *et al.,* 2018). The trimeric 3-fold symmetry and the 2-fold symmetry of the dimer would form a so-called ''hexagonal mesh'' which would be a platform for the anchoring function in the postsynaptic membrane (Figure 8). This theory was supported by a SAXs study, showing a relaxed and compact form of gephyrin, however, there are no atomic or near-atomic resolution as evidence for this theory (Sander *et al.,* 2003).

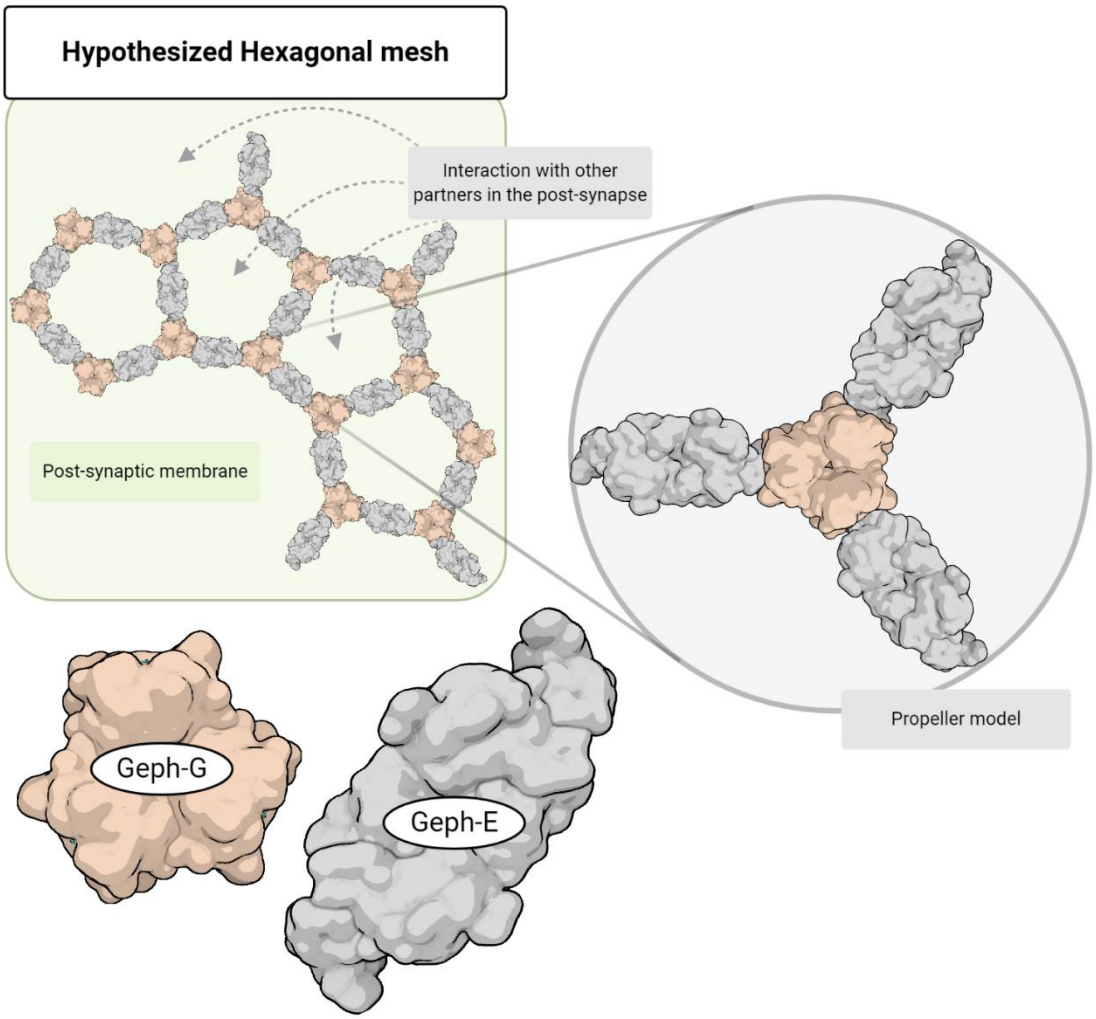



**Figure 8. The long-ongoing mesh hypothesis.** Since the discovery of gephyrin and the elucidation of the trimeric Geph-G (orange) and the dimeric Geph-E (grey) structures, a full picture of the complex assembly of gephyrin has never been shown. However, due to the 3-fold symmetry of the trimer and the 2-fold symmetry of the dimer, a hexagonal mesh arrangement hypothesis was proposed or the ''propeller model'' (Kneussel & Betz, 2000). The composition of the mesh would be seeded by a trimer, in which each of its monomers is connected to one monomer of the dimer, the other monomer of the dimmer is then connected to another monomer of another trimer (shown on the right). The final hexagonal mesh arrangement would form as a platform for the interaction of all the proteins in the post-synapse as well as an anchor for the receptors.

## 1.3. Protein complex assembly and evolution

### 1.3.1. Overview on protein complex assemblies

Proteins are continuously in physical contact with other proteins within the crowded intracellular environment. In a dynamic cellular context, protein complexes are molecular machineries that execute a vast range of biological functions that is crucial for organisms (Ellis, 2001). Many of the protein complexes are known to have a defined molecular function, which is understood based on a well-experimental characterization, conversely, there are plenty of complexes that are functionally and/or structurally unknown. This is due to the fact that these complexes are possibly not stable or transient (Landry *et al.,* 2013). Additionally, our techniques are not sophisticated enough to let us dig deeper into a dynamic living cell and have a complete picture of how protein complexes are assembled. Our understanding of protein complexes relies primarily on our techniques and, additionally, on the nature of the protein-protein interaction within the complex (Levy *et al.,* 2012). There is immense diversity when it comes to the specificity, rate, and duration of these interactions, as well as the nature of protein assemblies (Ellis, 2001).

### 1.3.2. Evolutionary aspect

The cell is full of short-lived interaction with unknown or little biological importance, but also with highly conserved protein assemblies that develop gradually throughout the evolutionary history of complexes. The resemblance of sequence identity usually correlates to a similar protein assembly (Landry *et al.,* 2013; Perkins *et al.,* 2010). For instance, proteins with 90% or greater sequence identity showed to have nearly conserved quaternary structures. However, 30 to 40 % identity of sequences decreases the likelihood of conservation of quaternary structures by 30% (Levy *et al.,* 2008). Moreover, the stoichiometry of protein complex assembly is evolutionary in constant alteration. Proteins tend to evolve based on the high demand for natural-selection and the need for new functional roles. Preferred self-assembly of the same symmetric subunits of a complex appears to be the common evolutionary transition (Perkins *et al.,* 2010). For example, a

monomeric protein forming a dimeric or cyclic $C_2$ or $C_n$ symmetry can evolve into a dihedral $D_n$.

On the other hand, the uneven stoichiometry of complexes is observed due to the existence of the same subunits in different environments. Strikingly, uneven stoichiometry across evolution is more enriched for bacterial complexes compared with eukaryotes (Marsh *et al.* 2014). Additionally, the uneven symmetry of protein complexes is shown to be observed more frequently *in vivo* than in the crystal structure (Marsh and Teichmann, 2014).

Interestingly, throughout the course of evolution, protein complexes tend to gain or lose protein subunits (Hashimoto and Panchenko, 2010). For instance, one of the prominent examples, the bacterial NADH-ubiquinone oxidoreductase (Complex I) has 14 distinct subunits, however, the same complex in the mammalian mitochondria has 46. An important example is the Mo-insertases, which is the focus of this work. The evolutionary Mo-insertase domain in prokaryotes is separately expressed. Conversely, later on, in the course of natural selection, the individual genes are fused in eukaryotes (Mendel, 2013). Accordingly, before forming complicated and multi-functioning complex structures in eukaryotes, a simplified Mo-insertase complex was formed in bacteria that is only responsible for Mo insertion (Xiang *et al.,* 2001; Krausze *et al.,* 2018). On the other hand, the splitting of genes into two or the reverse of gene fusion is less likely to occur than gene fusion. In general, the tendency of conserved interactions in different species is high. Almost all of the subunits of the protein complexes in yeast tend to be co-purified in comparison with their orthologs of human proteins (Krogan *et al.,* 2006; Van Dam & Snel, 2008).

## 1.3.3. Structural characterization of quaternary structures

Quaternary structures are the architectural assembly of which different distinct protein chains (subunits or tertiary structures) of a complex are structured according to one another (Klotz *et al.,* 1970). Quaternary structures can be homomeric, repeated copies of the same subunit for self-assembly, or heteromeric which is composed of different distinct protein subunits. In order to characterize these structures, different techniques could be implemented, including x-ray crystallography, nuclear magnetic resonance (NMR), electron microscopy (EM) or cryo-electron microscopy (Cryo-EM), Crosslinking Mass Spectrometry (XL-MS), native mass spectrometry (Native MS), small angle x-ray scattering (SAXs) and

computational modelling as well as others (Marsh and Teichmann, 2015) (Figure 9). Each technique has its advantages and disadvantages, nevertheless, combining various methods could lead to overcoming the drawbacks of the individual techniques. For example, it is not very reliable to get precise structural data from SAXs, as a standalone. On the other hand, combining XL-MS and Native MS with SAXs can generate accurate structural data (Blanchet & Svergun, 2013; Sinz *et al.*, 2015). There are many examples of combining techniques to shed light on protein complex assemblies. For instance, recently, the structure of the Neurotensin receptor 1 was prevailed in complex with β-arrestin using a combination of cryo-EM, XL-MS, as well as other additional methods and the same regarding the nuclear pore complex (Huang *et al.,* 2020; Bui *et al.,* 2013). Moreover, the structure of the tumour suppressor p53 was probed by XL-MS, native-MS and SAXs (Arlt *et al.,* 2015). Nevertheless, there is still technical progress in each of these methods in order to overcome their weaknesses.

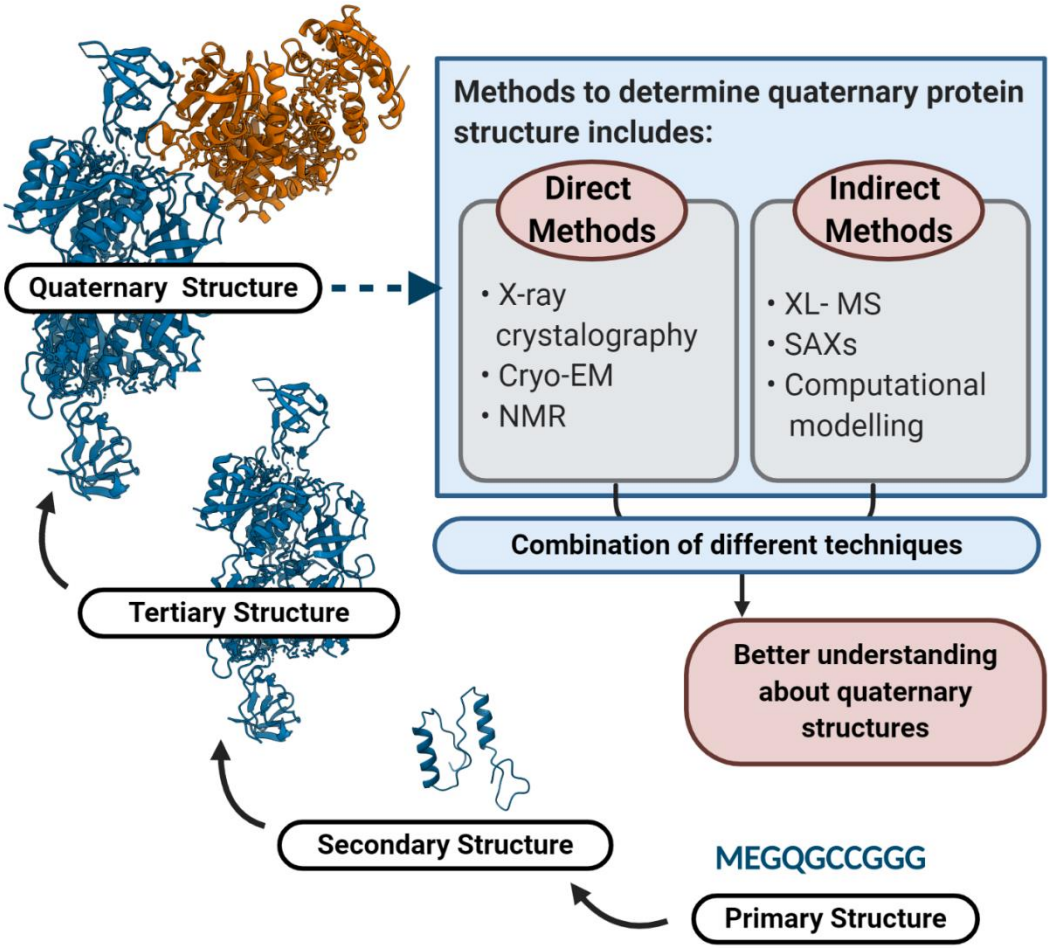



**Figure 9. Methods to determine the quaternary structure of proteins.** The amino acid sequence is referred to as the primary structure, and the secondary structure is an interaction between sections of a polypeptide chain, which includes α-helix and β-pleated sheet structures. Tertiary structure is a three-dimensional folding, while the quaternary structure is an alignment of subunits or multi-subunits to form small or large protein complexes. There are different methods in order to determine the quaternary structure, and these include direct and indirect methods. A combination of the different techniques is the most favoured option in order to have a better understanding of these protein complexes.

## 1.3.3.1. Crosslinking mass spectrometry (XL-MS)

To gain insight on conformation of a protein complex, crosslinking can be employed in order to covalently stabilize interaction partners *in vivo*. Chemical crosslinking relies on a covalent bond introduction between two functional groups of amino acids of different interaction partners or within the protein by a chemical substance (Sinz, 2006). Subsequently, the proteins of interest are enzymatically digested, and the peptides resulting are analyzed by

mass spectrometry with different approaches. The cross-linkers used are usually bi-functional with the same ''homobifunctional'' or different ''heterobifunctional'' reactive groups with a specified linker length range) (Sinz *et al.*, 2015). Upon analysis of the crosslinked products, the cross-linkers form ''molecular rulers'' in order to know more detailed structural insights of amino acids close in space, not in sequence (Mintseris & Gygi, 2020. Improvements in the mass spectrometers, enrichment methods, analysis methods, and the cross-linker reagents served in a more convenient as well as accurate identification of protein complexes with minimal false negatives (Bullock *et al.,* 2018; Sinz, 2006). Therefore, due to the continues advancements in XL-MS, many studies showed details about protein complex structures using chemical crosslinking as an additional or standalone tool.

### 1.3.3.2. Cryo-electron microscopy

In order to obtain information about macromolecular structures, cryo-electron microscopy (Cryo-EM) has been used, which relies on a particle beam of electrons shot at a frozen sample that illuminates a specimen to create a highly magnified image. The process starts with preparation of a homogenous protein sample, applied to a carbon grid, which is shock-freezed, and then the grid is inserted in an electron microscope. Subsequently, images are taken of 2D projections and particles are picked, aligned and averaged in order to create a 3D map. The map is processed into a 3D model of the protein structure (Renaud *et al.*, 2018; Murata & Wolf, 2018). For Cryo-EM, it is problematic to obtain structural data of complexes in high resolution (less than 2 or 1 Å). Furthermore, big complexes (>150 kDa) are more favored than smaller complexes (<150 kDa) and only in a near-atomic resolution (Lyumkis, 2019). Until recently, for the first time, method optimization of the EM lead to a Cryo-EM structure of β3 GABAA receptor homopentamer and Apoferritin at an atomic resolution of 1.7 Å and 1.22 Å, respectively (Nakane *et al.,* 2020*).*

### 1.3.3.3. Computational modelling

There is a large influx of near-atomic and atomic resolution structures of protein complexes using the previously discussed methods, however, vast numbers of complexes are still ambiguous due to the complexity and flexibility of such protein complexes (Marsh and Teichmann, 2015). Computational methods are an alternative in order to predict and

visualize protein complexes on a molecular level. There are different types of computational modelling, for example, template-based modelling and protein-protein docking (Wells *et al.,* 2018). The results from modelling are usually either confirmed, disregarded, or altered by experimental methods. Many protein structures turned out to be almost identical to the predicted models using different modelling softwares such as Rosetta, Modeller, iTasser (Simons *et al.*, 1999; Webb & Sali. 2016; Roy *et al.,* 2010). Some modellers demonstrated to be advantageous in *ab initio* modelling and protein-protein docking specifically, such as HADDOCK, which will be further discussed in this work. Massive improvement has recently been made with the computational prediction of protein structures (Koukos *et al.*, 2020). Fascinatingly, by deep learning, a group was capable of training a neural network in order to predict accurate information about structures and contact pairs of amino acids with 0.7 scores or higher, compared to a template (Senior *et al.,* 2020).

## 1.4. Outstanding questions (gaps of knowledge)

One of the significant gaps of knowledge that we have is due to the fact that there is no structure of the full-length gephyrin and/or cnx1. Hence, it is impossible to deduce: what is the architecture of the Mo-insertase complex? Is there a generic complex, or are there other possible conformations, and therefore, are the complexes homologues across the different species?

Another enigmatic question that evolves due to our lack of knowledge is; what is the exact domain conformation upon the product/substrate shuffling? Which amino acids are of absolute necessity for complex formation and for the transfer specifically? Is the complex present in the absence of the substrate and/or product? Furthermore, there is no evidence on direct interaction between Mo-insertase and the user enzyme.

## 1.5. Research aims of this work

This thesis aims to describe Mo-insertase complexes from different species, including *Escherichia coli, Arabidopsis thaliana,* and *Homo sapiens*, focusing on the conserved binding sites of Moco and its intermediates. Essentially, elucidating the structural design of the complex would explain the ''handshake'' that happens between the G- and E- domain as well as understanding the stoichiometry, which would shed light on the bigger picture of both cnx1 and gephyrin oligomerization. To that end, this thesis will focus on amino acid residues that impact complex formation in addition to the metabolic transfer. Lastly, the interaction of user enzymes with the Mo-insertases will be inspected.

# Chapter two

# 2. Materials and Methods

## 2.1. Materials

### 2.1.1. Equipment and instrumentation

**Table 1. All equipment and instrumentation used in this work.** The table lists the name of the device, model name and company.

| General name of the device | Model name | Company |
|---|---|---|
| Anion exchange chromatography column | SOURCE™ 15Q | GE Healthcare Life Sciences |
| Autoclave | FV13/11 proface | Ibs tecno mara |
| Balance | Precision balance 572 | KERN |
| Balance | Sartorius portable | SARTORIUS (FISCHER- WAA GEN) |
| Balance | Sartorius Analytics | SARTORIUS (FISCHER- WAA GEN) |
| Bench top centrifuge | My FUGE mini | Benchmark |
| Bucket for harvesting rotor | 1000 ml buckets | Beckman Coulter |
| Cell lysis system (Fast prep) | FastPrep-24™ | MP Biomedicals |
| Cell lysis system (French Press) | FRENCH® pressure cell press | SLM Aminco® |
| Cell lysis system (homogenizer) | Emulsiflex-C5 | AVESTIN |
| Centrifuge | Heraeus™ Pico™ 17 | Thermo Fisher Scientific |
| Centrifuge | Heraeus™ Fresco™ 21 | Thermo Fisher Scientific |
| Centrifuge for falcons and concentrators | Heraeus™ Multifuge™ X3R | Thermo Fisher Scientific |
| Centrifuge for harvesting | Avanti J-25 | Beckman Coulter |

| | | |
|---|---|---|
| Chromatography fractionator for gel filtration and ion exchange | Frac 920 | Amersham Pharmacia Biotech |
| Chromatography system for gel filtration and ion exchange | ÄKTA basic | Amersham Pharmacia Biotech |
| Chromatography system for gel filtration and ion exchange | ÄKTA purifier | Amersham Pharmacia Biotech |
| Chromatography system for HPLC | 1100 Series | Agilent |
| Freezer -20 | 230P-20 | LIEBHERR |
| Freezer -80 | MDF-U5386S | SANYO |
| Gel electrophoresis gel imagining system | ChemiDoc™ XRS+ | Bio-Rad |
| Gel electrophoresis power supply for SDS gels | PHERO-stab.300 | BIOTEC fischer |
| Gel electrophoresis power supply for SDS gels | PowerPac HC | Bio-Rad |
| Gel electrophoresis system for Agarose gels | Wide Mini-Sub Cell GT Cell | Bio-Rad |
| Gel electrophoresis system for Agarose gels | Mini-Sub Cell GT Cell | Bio-Rad |
| Gel electrophoresis system for SDS gels | Julaloo F10 | Julaloo |
| Gel electrophoresis system for SDS gels | Mini-PROTEAN® Tetra Vertical Electrophoresis Cell | Bio-Rad |
| Gel filtration column | Superdex 200 Increase 10/300 GL | GE Healthcare Life Sciences |
| Gel filtration column | Superdex 75 Increase 5/150 GL | GE Healthcare Life Sciences |
| Gel filtration column | Superdex 200 10/300 GL | GE Healthcare Life Sciences |
| HPLC column | ReproSil-Pur Basic-C18-HD | TECHLAB Gmbh |
| Ice machine | CoolNat | ZIEGRA |
| Incubator | Ecotron | INFORS HT |
| Incubator | Multitron standard | INFORS HT |
| Incubator | BODE | BODE |
| Laminar Flow Hood | HLB 2472 | Kendro |

| | | |
|---|---|---|
| Magnetic stirrer | RET basic | JK IKA labortechnik |
| Magnetic stirrer | Magnetrührwerk MOBIL 200 | VARIOMAG |
| Mass spectrometer for LC MS/MS | Nano-HPLC system coupled to a Nanospray Flex Ion Source of an Orbitrap Fusion Tribrid mass spectrometer | Thermo Fisher Scientific |
| Mass spectrometer for Native MS | Modified high-mass QToF II instrument | Waters Micromass/MS Vision |
| Microplate Spectrophotometer | Multiskan go thermo | Thermo Fisher Scientific |
| Microwave | Klarstein MaxiWave | Klarstein |
| PCR cycler | Life ECO | BIO ER |
| PCR cycler | Flex cycler | Analytik Jena |
| pH meter | ph-meter 763 Multi-calimotrices | Knick |
| Pipettes | Pipetman | Gilson |
| Pipettes | Pipet-lite XLS | Rainin |
| Protein storage | LS 750 | TechLab Gmbh |
| Refrigerator 4° C | IK 1920-20 | LIEBHERR |
| Rotor for falcons and concentrators | Fiberlite™ F15-8 x 50cy | Thermo Fisher Scientific |
| Rotor for harvesting | JLA-9.1000 | Beckman Coulter |
| Shaker | W55 | Edmund Bühler |
| Shaker | 3015 GFL | Wilh. O. Schmidt Gmbh |
| Sonication bath | Bandelin Sonorex | Reichmann |
| Talos electron microscope | FC200 | Thermo Fisher Scientific |
| Thermoblock | ThermoStat Plus | Eppendorf |
| Thermomixer | Thermomixer C | Eppendorf |
| Thermomixer | Thermomixer comfort | Eppendorf |
| Transfer system for Western blotting | Trans-Blot Turbo | Bio-Rad |
| UV/Vis Spectrophotometer | Ultrospec™ 2100 pro | Amersham Bioscience |
| UV/Vis Spectrophotometer | Lambda 25 | PerkinElmer |
| Vacuum pump system | Membran-Vakuumpumpe | VACUUBRAND Gmbh |

| | | |
|---|---|---|
| Vortex mixer | Vortex GENEIE 2 | Scientific industries |
| Water bath | Mini therm1 | DINKELBERG |

## 2.1.2. Kits

**Table 2. All kits used in this work.** The table includes the name of the kit used, the purpose of use and the company supplying it.

| Name of the kit | Purpose of use | Company |
|---|---|---|
| BSA standard kit | Standard for concentration determination | Thermo Fisher Scientific |
| Gel Filtration Calibration Kit HMW | Protein standard for the calibration for gel filtration columns | GE Healthcare Life Sciences |
| Kit for Molecular Weights 29,000 – 700, 000 For Gel Filtration Chromatography | Protein standard for the calibration for gel filtration columns | Sigma Aldrich |
| Monarch® DNA Gel Extraction Kit | Isolation of DNA from agarose gel | New England Biolabs |
| Monarch® Plasmid Miniprep Kit | Plasmid DNA isolation | New England Biolabs |
| NEBuilder® HiFi DNA Assembly Cloning Kit | Seamless cloning | New England Biolabs |
| Q5® Site-Directed Mutagenesis Kit | Amino acid site directed mutagenesis | New England Biolabs |
| SERVALight Eos | Visualization of western blot | SERVA |
| TGX Stain-Free™ FastCast™ Acrylamide Solutions | Casting SDS PAGE gels | Bio-Rad |
| Trans-Blot® Turbo™ Transfer | Western blotting of SDS PAGE gels | Bio-Rad |

## 2.1.3. Consumables

**Table 3 All consumables used in this work.** The table includes the material used and the company supplying it.

| Material used | Company |
|---|---|
| 1.5 ml tubes | SARSTEDT |
| 10 µl pipette tips | SARSTEDT |
| 10 ml columns | Bio-Rad |
| 1000 µl pipette tips | SARSTEDT |
| 14 ml tubs | SARSTEDT |
| 15 ml tubes | SARSTEDT |
| 2 ml tubes | SARSTEDT |
| 200 µl pipette tips | SARSTEDT |
| 25 ml columns | Bio-Rad |
| 5 ml pipette tips | SARSTEDT |
| 50 ml syringes | VWR |
| 50 ml tubes | SARSTEDT |
| Cryo micro tubes | SARSTEDT |
| Cuvettes | SARSTEDT |
| Disposable graduated transfer pipets | VWR |
| Eppendorf conical tubes | Eppendorf |
| KIMTECH | Kimberly-clark professional |
| Microtest plate 96 well | SARSTEDT |
| Multiply® PCR Tubes | SARSTEDT |
| Parafilm | Curwood/ Fisher scientific |
| PD10 columns | GE Healthcare Life Sciences |
| Petridish 92 x 16 mm | SARSTEDT |
| Rotilabo® Aluminumfoile 30 µm | Carl Roth |
| Roti-protect nitrite gloves | Carl Roth |
| S1 Disposable bags | SARSTEDT |
| Syringe filters, Filtropur S 0.2 | SARSTEDT |
| Transfer pipettes | SARSTEDT |

| | |
|---|---|
| Vivaspin® 15 10,000 MWCO | Sartorius |
| Vivaspin® Turbo 4 10,000 MWCO | Sartorius |
| Vivaspin® Turbo 4 30,000 MWCO | Sartorius |

## 2.1.4.Chemicals

**Table 4. All chemicals used in this work.** The table includes the chemical name and the company supplying it.

| Chemical | Company |
|---|---|
| 1,5-difluoro-2,4-dinitrobenzene (DNDFB) | Thermo Fisher Scientific |
| 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) | TCI |
| 2- propanol | Carl Roth |
| 2-mercaptoethanol | Sigma-Aldrich |
| Acetic acid | Merck |
| Acetone | VWR |
| Agarose standard | Carl Roth |
| Alkaline phosphatase | New England Biolabs/Roche |
| Ammonium persulfate | SERVA |
| Ascorbic acid | Sigma-Aldrich |
| Bacillol AF | HARTMANN |
| Bis-Tris | Carl Roth |
| Bromophenol Blue | Merck |
| cOMPLETE tablets | Roche |
| D-desthiobiotin | IBA |
| Dimethyl sulfoxide | Sigma-Aldrich |
| Disuccinimidyl dibutyric urea (DSBU) | Thermo Fisher Scientific |
| Disuccinimidyl suberate (DSS) | TCI |
| Disuccinimidyl sulfoxide (DSSO) | Thermo Fisher Scientific |

| | |
|---|---|
| Disuccinimidyl tartrate (DST) | Thermo Fisher Scientific |
| DNase I | Roche |
| EDTA | SERVA |
| Ethanol | Sigma-Aldrich |
| Formaldehyde | Merck |
| Glycerin | Carl Roth |
| Guanidinium hydrochloride | SERVA |
| HEPES | SERVA |
| Hydrochloric acid 37% | Merck |
| Imidazole | Sigma-Aldrich/Merck |
| Isopropyl-β-D-thiogalactopyranosid | SERVA |
| Methanol | VWR |
| Methanol HPLC grade | Fischer scientific |
| Nickel sulfate | Sigma-Aldrich |
| Ni-NTA agarose | Qiagen |
| Phosphodiesterase | Abnova |
| ROTI®Quant | Carl Roth |
| Serva Blue G | SERVA |
| Serva Blue R | SERVA |
| Skimmed milk powder | Heirler |
| Sodium chloride | Carl Roth |
| Sodium dodecyl sulfate | SERVA |
| Sodium hydroxide | Sigma-Aldrich |
| Sodium Molybdate | Sigma-Aldrich |
| Strep-Tactin® Superflow® high capacity resin | IBA |
| Tris Hydrochloride | Duchefa Biochemie |
| Tris ultra-pure | Duchefa Biochemie |
| Triton-X | Carl Roth |
| Tryptone | Duchefa Biochemie |
| Tween 20 | Carl Roth |
| UREA | Duchefa Biochemie |

| | |
|---|---|
| Yeast extract | Duchefa Biochemie |

## 2.1.5. Sample loading buffers and ladders

**Table 5. All sample loading buffers and ladders used in this work.** The table includes the name of the buffer or ladder, the purpose of use and the company supplying it.

| Name | Purpose of use | Company |
|---|---|---|
| Color Prestained Protein Standard, Broad Range | Protein ladder | New England Biolabs |
| GeneRuler 100 bp | DNA ladder | Thermo Fisher Scientific |
| GeneRuler 1KB | DNA ladder | Thermo Fisher Scientific |
| Green loading dye | DNA loading & tracking dye | Thermo Fisher Scientific |
| Midori Green | In-gel staining of DNA agarose gel | NIPPON Genetics Europe |
| Prestained Ladder | Protein ladder | Thermo Fisher Scientific |
| Prestained Pageruler Protein Ladder | Protein ladder | Thermo Fisher Scientific |
| Purple loading dye | DNA loading & tracking dye | New England Biolabs |
| Spectra HR Ladder | Protein ladder | Thermo Fisher Scientific |
| Unstained Pageruler Protein Ladder | Protein ladder | Thermo Fisher Scientific |

## 2.1.6. Buffers and other solutions

**Table 6. All buffers and solutions used in this work.** The table includes buffer or solution name and composition.

| Buffer name | Composition |
|---|---|
| 5x loading buffer for SDS | 10% SDS, 500mM DTT, 50% Glycerol, 250mM Tris-HCl pH 6.8 and 0.5% bromophenol blue dye |
| De-staining solution for SDS | 50% (v/v) methanol in water with 10% (v/v) acetic acid |

| | |
|---|---|
| *E.coli* glycerol storage buffer | 65 % glycerin, 100 mM $MgSO_4$ and 25 mM Tris-HCl |
| Gel filtration buffer | 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA |
| Gel filtration buffer for crosslinking | 100 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM EDTA |
| Ion-exchange high salt buffer | 20 mM Tris-HCl pH 7 and 1 M NaCl |
| Ion-exchange low salt buffer | 20 mM Tris-HCl pH 7 |
| Native running buffer | 12.5 mM Tris pH 8-9, 96 mM glycine |
| Ni-NTA elution buffer | 100 mM HEPES, 300 mM NaCl, 250 mM Imidazol and 10 % (v/v) Glycerol |
| Ni-NTA lysis buffer | 100 mM HEPES, 300 mM NaCl, 10 mM Imidazol and 10 % (v/v) Glycerol |
| Ni-NTA regeneration 1 | 6 M GuHCl and 0.2 M acetic acid |
| Ni-NTA regeneration 2 | 100 mM $NiSO_4$ |
| Ni-NTA wash buffer | 100 mM HEPES, 300 mM NaCl, 20 mM Imidazol and 10 % (v/v) Glycerol |
| Strep-tactin elution buffer | 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 10 % (v/v) Glycerol and 5 mM Desthiobiotin |
| Strep-tactin elution buffer for crosslinking | 100 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM EDTA, 5 % (v/v) Glycerol and 5 mM Desthiobiotin |
| Strep-tactin lysis buffer | 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 10 mM EDTA, 10 % (v/v) Glycerol |
| Strep-tactin lysis buffer for crosslinking | 100 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM EDTA, 5 % (v/v) Glycerol |
| Strep-tactin wash buffer | 100 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM EDTA, 10 % (v/v) Glycerol |
| Strep-tactin wash buffer for crosslinking | 100 mM HEPES (pH 7.5), 300 mM NaCl, 1 mM EDTA, 5 % (v/v) Glycerol |
| TAE buffer | 2 M Tris-HCl pH 8.0, 1 M acetic acid and 14.6 mg/ml EDTA |
| TBS (10x) | 24 mg/ml Tris pH 7.6 and 88 mg/ml NaCl |
| TBS-T (50x) | 100 ml of TBS (10x) with 4 ml of 25 % Tween20 per 1 liter |
| Transfer buffer | 25 mM Tris, 192 mM glycine pH 8.3. |

## 2.1.7. Oligonucleotides

### 2.1.7.1. Primers

All primers were purchased from Sigma Aldrich/ Merch. Primers were either manually prepared, and $T_m$ was measured using NEB primer design online tool OR using SnapGene (GSL Biotech LLC).

**Table 7. All primers used for this work.** The table includes the name of the primer and the sequence

| Primer name | Sequence | Purpose of use |
|---|---|---|
| Build_Cnx1E_D088A_For | CACTCCAAGAGCGGTAACACCTGAAGCAACGAAAAAGG | Generation of D548A mutant |
| Build_Cnx1E_D088A_Rev | GTGTTACCGCTCTTGGAGTGAAGCCAGTAC | |
| Build_Cnx1E_E040A_K041A_for | TCCTCGTCTGCGGCGCTGGGGGGAGCTAAGGTAG | Generation of E500A, K501A mutant |
| Build_Cnx1E_E040A_K041A_rev | CCCAGCGCCGCAGACGAGGAATCAACTAC | |
| Build_Cnx1E_E162A_K163A_For | TATTGCGGCGGATGCAACTGTACTAACAACTGGAG | Generation of E157A, K158A mutant |
| Build_Cnx1E_E162A_K163A_Rev | CAGTTGCATCCGCCGCAATATCACAACCCACTCTAC | |
| Build_Cnx1E_K149A_K150A_For | AACCGCGGCCGGCACCGATATCCGTAGAGTG | Generation of K144A, K145A mutant |
| Build_Cnx1E_K149A_K150A_rev | TATCGGTGCCGGCCGCGGTTTGAATCAGTATTTTTACTC | |
| built_hs_geph_P2_for | CAGCAGCTATTGCTGCAAAGATTCCAG | Generation of *Homo sapien* gephyrin isoform 2 |
| built_hs_geph_P2_rev | GCAGCAATAGCTGCTGCAGTAATATG | |
| Cnx1E_D160A_for | CTATTGAGAAAGATGCAACTGTAC | Generation of D155A mutant |
| Cnx1E_D160A_rev | CACAACCCACTCTACGGA | |
| Cnx1E_R155A_for | GCAAGAGTGGGTTGTGATATTG | Generation of R150A mutant |

| | | |
|---|---|---|
| Cnx1E_R155A_rev | GATATCGGTGCCTTTCTTG | |
| Cnx1E_R156A_for | GCAGTGGGTTGTGATATTGAGAAAG | Generation of R151A mutant |
| Cnx1E_R156A_rev | ACGGATATCGGTGCCTTT | |
| Cnx1E_R407A_for | GCGCTTTTAAGTATGAGGTCG | Generation of R402A mutant, generation of R402A, E607A mutant, and generation of R402A, E610A mutant |
| Cnx1E_R407A_rev | ACTGCTCATCTGATGTCC | |
| Cnx1FL_S269D_D274S_for | CATTTCGTCAAGCCATTGCTC | Generation of S269D, D274S mutant in Cnx1 |
| Cnx1FL_S269D_D274S_rev | CCTGTCTCCCATATCAAC | |
| Cnx1FL_S269D_for | GATATGGGAGACAGGGATTTC | Generation of S269D mutant in Cnx1 |
| Cnx1FL_S269D_rev | AACACCACCAGAAGTTAAAATAATATC | |
| Cnx1G_E152A_for | CGTGTATGGAGGCTCTGTTACCTG | Generation of E607A mutant and generation of R402A, E607A mutant |
| Cnx1G_E152A_rev | CGGCCACTGCGTTTGGGT | |
| Cnx1G_E155A_for | CGGCTCTGTTACCTGCATTG | Generation of E610A mutant and generation of R402A, E610A mutant |
| Cnx1G_E155A_rev | CCATACACTCGGCCACTG | |
| Cnx1G_N146A_for | GCCCCAAACGCAGTGGCCGA | Generation of N601A mutant |
| Cnx1G_N146A_rev | TCCAGGCATGTTGATTATCAATGTTGAG | |
| Cnx1G_N148A_for | GCCGCAGTGGCCGAGTGTAT | Generation of N603A mutant |
| Cnx1G_N148A_rev | TGGGTTTCCAGGCATGTTG | |
| F_470_Cnx1G_for | TGAAGCGCTATTAAAGAGGAGAAATTAACTATGGT | Generation of 7saltbridges construct and generation of 7saltbridges+3charged residues construct |
| F_470_Cnx1G_rev | CTTAGCGCTTTAGTGAGTACAAGAACATCC | |

| | | |
|---|---|---|
| Primer 1614 Split Amp rev | GAGCACTTTTAAAGTTCTGCTATGTGGC | Gibson assembly |
| Primer 1621 Split Amp for | GCAGAACTTTAAAAGTGCTCATCATTGGAAA | |
| V_Cnx1G_470_for | ACTCACTAAAGCGCTAAGCTTAATTAGCT | Generation of 7saltbridges construct and generation of 7saltbridges+3charged residues construct |
| V_Cnx1G_470_rev | CCTCTTTAATAGCGCTTCATTACTTCTCG | |
| QE-for | GTATCACGAGGCCCTTTCG | for sequencing |
| QE-rev | GTTCTGAGGTCATTACTGG | |

## 2.1.7.2. Gene synthesis

**Table 8. Gene synthesized for this work.** The table includes the name and sequence of the constructs.

| **7 salt bridges construct was ordered from BioCat with the sequence:** |
|---|
| AGCGCTATTAAAGAGGAGAAATTAACTATGGTACCTGGGCCTGAGTATAAAGTGGCTATTCTTACCGTCAGTGAT<br>ACTGTCTCAGCTGGGGCTGGACCTGATCGAAGTGGGCCCAGAGCTGTGTCAGTAGTTGATTCCTCGTCTGAAAA<br>ACTGGGGGGAGCTAAGGTAGTTGCAACAGCTGTTGTCCCCGATGAAGTGTCCAGAATATCGTCTATTCTTCAGA<br>AATGGAGCGATGTTGACGAAATGGATCTCATTCTTACACTTGGTGGTACTGGCTTCACTCCAAGCGATGTAACAC<br>CTGAAGCAACGAAAAGCGTGATCAGCAGAGAAACACCCGGTCTCCTCTTTGTTATGATGCAAGAGAGCTTAAAG<br>ATCACTCCATTTGCGATGCTCTCGCGCTCTGCGGCAGGAATTAGAGGCTCAACATTGATAATCAACATGCCTGGA<br>AACCCAAACGCAGTGGCCGAGTGTATGGAGGCTCTGTTACCTGCATTGAAACACGCACTGAAGCAGATCAAAGG<br>AGACAAGAGAGAGAAACATCCGAAGCACATACCTCACGCAGAAGCCACCCTACCGACTGATACATGGGACCAAA<br>GCTATAAGTCAGCCTATGAGACGGGAGAGAAGAAAGAAGAGGCTGGATGTTCTTGTACTCACTAAAGCGCT |
| |
| **7 salt bridges construct + 3 charged residues was ordered from BioCat with the sequence:** |
| AGCGCTATTAAAGAGGAGAAATTAACTATGGTACCTGGGCCTGAGTATAAAGTGGCTATTCTTACCGTCAGTGAT<br>ACTGTCTCAGCTGGGGCTGGACCTGATCGAAGTGGGCCCAGAGCTGTGTCAGTAGTTGATTCCTCGTCTGAAAA<br>ACTGGGGGGAGCTAAGGTAGTTGCAACAGCTGTTGTCCCCGATGAAGTGTCCAGAATATCGCGCATTCTTCAGA<br>AATGGAGCGATGTTGACGAAATGGATCTCATTCTTACACTTGGTGGTACTGGCTTCACTCCAGAAGATGTAACAC<br>CTGAAGCAACGAAAGATGTGATCAGCAGAGAAACACCCGGTCTCCTCTTTGTTATGATGCAAGAGAGCTTAAAG<br>ATCACTCCATTTGCGATGCTCTCGCGCTCTGCGGCAGGAATTAGAGGCTCAACATTGATAATCAACATGCCTGGA<br>AACCCAAACGCAGTGGCCGAGTGTATGGAGGCTCTGTTACCTGCATTGAAACACGCACTGAAGCAGATCAAAGG |

| |
|---|
| AGACAAGAGAGAGAAACATCCGAAGCACATACCTCACGCAGAAGCCACCCTACCGACTGATACATGGGACCAAA GCTATAAGTCAGCCTATGAGACGGGAGAGAAGAAAGAAGAGGCTGGATGTTCTTGTACTCACTAAAGCGCT |

## 2.1.8. Antibiotics, lac-operon inductor and Medium

For all constructs, ampicillin (Carl Roth/ SERVA) used was in 100 mg/ml concentration obtained from Carl Roth. Unless stated otherwise, 1 mM final concentration of Isopropyl-β-D-thiogalactopyranosid (IPTG) was used as lac-operon inducer. The medium type and preparation components used are shown in the table below (Table 9).

**Table 9. All medium used in this work.** The table contains the names and components of all medium prepared

| | 2YT | LB | SOC |
|---|---|---|---|
| **NaCl** | 5 g/L | 10 g/L | 0.5 g/L |
| **Other components** | None | None | 2.5 mM KCl<br>10 mM MgCl2<br>10 mM MgSO4<br>20 mM Glucose |
| **Tryptone/ Peptone** | 16 g/L | 10 g/L | 20 g/L |
| **Yeast exctract** | 10 g/L | 5 g/L | 5 g/L |

## 2.1.9. Antibodies

**Table 10. Antibodies used in this work.** The table includes names and references of the antibodies

| **Primary antibodies used for immunoblotting** |
|---|
| 7G8 - Strep-tag® II (Ringel *et al.,* 2015) |
| Anti-MoeA (generated in this work) |
| Anti-Nit-9 (Ringel *et al.*, 2013) |
| **Secondary antibody used for immunoblotting** |
| Peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG + IgM (H+L), Dianova |

## 2.1.10. *Escherichia Coli* strains

**Table 11. All strains used in this work.** The table includes strain name, genotype and source

| Strain Name | Genotype | Source |
|---|---|---|
| BL21(DE3) | F- ompT gal dcm lon hsdSB(rB -mB - ) λ(DE3 [lacI lacUV5-T7 gene1 ind1 sam7 nin5]) | Studier and Moffatt, 1986 |
| DH5α | supE 44Δlac U169 (Δ80lacZΔM15) hsdR17 recA endA1 gyrA96 thi-1 relA1 | Hanahan, 1983 |
| KJW2-A | Nonpolar Keio *E. coli* K-12 derivative: ΔMoaA | K. Warnhoff and G. Ruvkun, 2019 |
| KJW3-A | Nonpolar Keio *E. coli* K-12 derivative: ΔMoaC | K. Warnhoff and G. Ruvkun, 2019 |
| NEB® 10-beta | Δ(ara-leu) 7697 araD139 fhuA ΔlacX74 galK16 galE15 e14- φ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL (StrR) rph spoT1 Δ(mrr-hsdRMS-mcrBC) | New England Biolabs |
| NEB® 5-alpha | fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi1 hsdR17 | New England Biolabs |
| RK5204 | RK4353 *chlA204*::Mu cts | Studier and Moffatt, 1986 |
| RK5206 | RK4353 *chlA206*::Mu cts | Studier and Moffatt, 1986 |
| SE1581 | BW545 *moeA101* | Adnan Hasona et al., 1998 |
| TP1000 | F´ ΔlacU169 araD139 rpsL150 relA1 ptsF rbsR flbB Δ(mobAB) | Palmer et al., 1996 |

## 2.1.11. Vectors

**Table 12. All vectors used in this work.**
The table includes vector names.

| Vector |
|---|
| pTWO-C |
| pTWO-N |

| |
|---|
| pDT |
| pG+ |
| pQE80 |

## 2.1.12. Expression constructs

**Table 13. All constructs used for protein expression in this work.** The table contains construct name, insert name, vector and mutation

| Construct number | Insert name | Vector | Mutation (if present) |
|---|---|---|---|
| #470 | Cnx1E S269A D274A | pG+ | Cnx1E: S269A D274A |
| #471 | Cnx1E wt | pG+ | None |
| #509 | Cnx1FL wt | pDT | None |
| #510 | Cnx1FLS269D D274S | pDT | Cnx1E: S269D D274S |
| #557 | Nit9 | pTwo-C | None |
| #559 | Cnx1FL | pTwo-C | None |
| #561 | MoeA | pTwo-C | None |
| #562 | MogA | pTwo-C | None |
| #614 | MoeA-Cnx1G | pG+ | None |
| #615 | Cnx1E-MogA | pG+ | None |
| #619 | MoeAMogA | pG+ | None |
| #657 | StrepII-C | none | None |
| #658 | StrepII-N | none | None |
| #718 | MogA S117A | pEX-A128 | S117A |
| #721 | MoeA switch v2 | pEX-A 128 | None |
| #750 | Cnx1E S269D D274s Cnx1G E152 | pG+ | Cnx1: S269D D274S, E607A |
| #751 | Cnx1E S269D D27A R407A | pG+ | |
| #752 | Cnx1E S269D D27A Cnx1G E155A | pG+ | Cnx1: S269D D27A, E610A |

| | | | |
|---|---|---|---|
| #753 | Cnx1E S269D D27A D160A | pG+ | Cnx1: S269D D27A, D155A |
| #754 | Cnx1E S269D D27A Cnx1G N146A | pG+ | Cnx1: S269D D27A, N601A |
| #755 | Cnx1E S269D D27A R155A | pG+ | Cnx1: S269D D27A, R150A |
| #756 | Cnx1E S269D D27A Cnx1G E152A | pG+ | Cnx1: S269D D27A, E607A |
| #757 | Cnx1E S269D D27A Cnx1G E155A | pG+ | Cnx1: S269D D27A, E610A |
| #758 | Cnx1G 7salt bridges | pUC75 | Cnx1G: 6 mutations included in supplemental |
| #759 | Cnx1G 7salt bridges 3charges | pUC75 | Cnx1G: 6 mutations included in supplemental |
| #760 | Cnx1E S269D D27A Cnx1G N148A | pG+ | Cnx1: S269D D27A, N603A |
| #761 | Cnx1E S269D D27A Cnx1E R156A | pG+ | Cnx1: S269D D27A, R151A |
| #762 | Gephyrin rn P1 | pTWO-C | None |
| #763 | Gephyrin rn P2 | pTWO-C | None |
| #764 | Cnx1G 7salt bridges | pG+ | Cnx1G: 6 mutations included in supplemental |
| #765 | Cnx1G 7salt bridges 3charges | pG+ | Cnx1G: 6 mutations included in supplemental |
| #766 | Gephyrin hs P2 | pTWO-C | None |
| #767 | Cnx1E S269D D27A Cnx1G D088A | pG+ | Cnx1: S269D D27A, D548A |
| #768 | Cnx1E S269D D27A Cnx1G E040A K041A | pG+ | Cnx1: S269D D27A, E500A, K501A |
| #769 | Cnx1E S269D D27A K149A K150A | pG+ | Cnx1: S269D D27A, K144A, K145A |

| | | | |
|---|---|---|---|
| #770 | Cnx1E S269D D27A E162A K163A | pG+ | Cnx1: S269D D27A, E157A, K158A |

## 2.1.13. Enzymes

**Table 14. All enzymes used in this work.** The table includes the enzyme name, and the company supplied it.

| Enzyme name | Company |
|---|---|
| BamHI | Thermo Fisher Scientific/ New England Biolabs |
| Eco47III (AfeI) | Thermo Fisher Scientific/ New England Biolabs |
| HindIII | Thermo Fisher Scientific/ New England Biolabs |
| KLD enzyme mix | New England Biolabs |
| PstI | Thermo Fisher Scientific |
| Q5 high fidelity | New England Biolabs |
| SmaI | Thermo Fisher Scientific |
| Taq DNA polymerese | New England Biolabs |
| XhoI | Thermo Fisher Scientific |
| XmaI | Thermo Fisher Scientific |

## 2.1.14. Software and servers

**Table 15. All software and servers used in this work.** The table includes software name, company or source and purpose of use.

| Software name | Company and/or source | Purpose |
|---|---|---|
| ATLAS | Thermo Fisher Scientific | Negative staining and microscopy |
| autoPROC | Global Phasing Ltd | Data set processing for crystalization |
| BUSTER | Bricogne *et al.* 2017 | Struture refinement |

| | | |
|---|---|---|
| Chromas | Technelysium | Sanger sequencing analysis |
| ClusPro | Vajda Lab and ABC Group Boston University and Stony Brook University | Protein-protein docking |
| Cluster omega | European Molecular Biology Laboratory EMBL-EBI | Multiple sequence alignment |
| Conserved Domain | NCBI Shennan Lu *et al.* | Domain structure comparison |
| Coot | Paul Emsley (MRC-LMB at Cambridge) | Structural editing |
| Cport | Bonvin Lab | Generation of active and passive residues for modelling |
| CX-Circos | Chait Lab | Visualization of the cross-links |
| EPU theromos | Thermo Fisher Scientific | Negative staining and microscopy |
| ExPASy PeptideMass | Swiss Institute of Bioinformatics (SIB) | *In silico* digestion of proteins |
| ExPASy ProtParam | Swiss Institute of Bioinformatics (SIB) | Protein parameters |
| FiberDock | E. Mashiach, R. Nussinov and H. J. Wolfson | Backbone refinement in molecular docking |
| GraphPad Prism | GSL Biotech LLC | Analysis and graph presentation |
| HADDOCK | Bonvin Lab | Modelling/ protein-protein docking |
| Image Lab | Bio-Rad | Visualization and documentation of SDS PAGE and agarose gels |
| I-TASSER | Ambrish Roy , Alper Kucukural & Yang Zhang | Modelling |
| MeroX | Michael Gutze & Sinz Lab | Analysis of the Cross-links |
| MobaXterm | Mobatek | SSH terminal, telnet, EDP, VNC and Mosh for windows |
| Multialign | Florence Corpet | Sequence alignment |
| OpenMS | Strum M, et al., 2008 | MS results analysis |
| PDB Flex | GodzikLab | Protein flexibility analysis tool |

| | | |
|---|---|---|
| Phaser | McCoy *et al.,* 2007 | Solving the crystallographic phase problem |
| PHENIX | D. Liebschner *et. al* 2019 | Structural manipulation and refinements |
| Phyre2 | Structural Bioinformatics Group, Imperial College, London Lawrence Kelley, Michael Sternberg | Homology modelling |
| PISA | European Molecular Biology Laboratory EMBL-EBI | Protein interface exploration tool |
| Proteome Discoverer | Thermo Fisher Scientific | MS results analysis |
| PyMOL | Schrödinger, Inc. | Structural visualization and editing |
| RELION | Scheres 2012 | Image analysis, 2D and 3D classifications |
| Robetta | Baker Lab | *Ab initio* modelling |
| SnapGene | GSL Biotech LLC | DNA cloning, visualization and planning |
| STARANISO | Global Phasing Ltd | Anisotropy |
| TEM user software | Thermo Fisher Scientific | Negative staining and microscopy |
| UCSF Chimera | Resource for Biocomputing, Visualization, and Informatics (RBVI) – University of California | Structural visualization and editing |
| UCSF ChimeraX | Resource for Biocomputing, Visualization, and Informatics (RBVI) – University of California | Structural visualization and editing |
| Xcalibur | Thermo Fisher Scientific | Software for fusion/Orbitap |

## 2.2. Methods

### 2.2.1. DNA methods

## 2.2.1.1. Transformation of *Escherichia Coli*

10-100 ng of target DNA was added to 10 to 100 µl aliquots of chemically competent *E. coli* cells on ice for 30 min. Afterwards, the cells were heat-shocked for 30 to 60 sec (30 sec for NEB®5α and NEB®10β, while 60 sec for BL21 (DE3), RK5204, RK5206, TP1000 and SE1581) in a Thermomixer (Eppendorf). The cells were then incubated on ice for 3 min, and 900 µl of LB or SOC medium were added. Afterwards, the cells were shaken for 60 min in Ecotron incubator (INFRON HF) at 37°C, for NEB®5α, TP1000, SE1581 and NEB®10β, and at 28°C, for RK5204 and RK5206 at 130 - 200 rpm. Subsequently, the cells were applied on an LB-agar plate with 100 mg/ml ampicillin (Carl Roth/SERVA) and incubated at 37°C or 28°C overnight. The plates were wrapped in parafilm and stored at 4°C for further cloning or expression usage.

**Table 16. Strains used for transformation.** The table includes the strain names, temperature and heat shock time required

| | NEB®5α | NEB®10β | BL21 (DE3) | TP1000 | SE1581 | RK5204 | RK5206 |
|---|---|---|---|---|---|---|---|
| **Temperature** | 37°C | 37°C | 37°C | 37°C | 37°C | 28°C | 28°C |
| **Heat Shock time** | 30 sec | 30 sec | 50 - 60 sec | 50 - 60 sec | 50 - 60 sec | 50 - 60 sec | 50 - 60 sec |

## 2.2.1.2. Cultivation of *Escherichia Coli*

The *E. coli* colonies were grown in Lysogeny Broth (LB), 2YT or SOC medium (New England Biolabs) at 37 °C or 28 °C depending on the strain (see 2.1.10). For recombinant gene expression or to obtain more DNA, single colonies were picked and transferred into tubes with 2 – 5 ml of medium, including 100 mg/ml ampicillin. The liquid medium was shaken at 130 – 200 rpm for 12 – 16 hrs. To prepare LB solid medium with 100 mg/ml ampicillin, 1.5% (w/v) agar-agar (COMPANY) and places were stored routinely at 4 °C.

### 2.2.1.3. Storage of *Escherichia Coli*

In order to prepare glycerol stocks, 500 µl of a solution of 65% (v/v) glycerol,100 mM $MgSO_4$ and 25 mM Tris-HCl pH 8.0 was used with an equal amount of 500 µl of overnight culture, which was shock frozen in liquid nitrogen and kept at in – 80 °C Freezer.

### 2.2.1.4. DNA concentration determination

In order to assess the quality and concentration of DNA, Multiskan go thermos (Thermo Fischer Scientific) was used. Primarily, concentration measurement was done by measuring UV-VIS absorbance at 260 nm ($A_{260}$). Other contaminants were checked for by measuring absorbance from 230 to 320 nm. The $A_{260}/A_{280}$ ratio in the range of 1.7 to 2.0 was further used. Secondarily, another method was used by comparing the DNA loaded on an agarose gel with a DNA quantitation standard or ladder.

### 2.2.1.5. DNA gel electrophoresis

Agarose gel electrophoresis was used for the assessment of the quality and size of DNA. The agarose gel contained 0.5 to 1% (w/v) agarose standard (Carl Roth) in 1x TAE buffer. For in-gel staining, 3 µl of Midori Green (NIPPON Genetics Europe) per 50 mL gel. The DNA samples were mixed with purple (New England Biolabs) or green (Thermo Fischer Science) loading buffer and loaded on the casted agarose gel in the Wide or Mini-Sub Cell GT Cell (Bio-Rad). A DNA ladder, 1kb or 100 bp ladder (Thermo Fischer Scientific), was used as a standard to size the DNA. The gel was subjected to gel electrophoresis in 1x TAE buffer for a duration of 30 – 45 min at a voltage of 120 V. ChemiDoc™ XRS+ system (Bio-Rad) was used to visualize the gel using ImageLab software (Bio-Rad). The gel was visualized using Midori Green which incorporated into the individual DNA fragments.

### 2.2.1.6. Plasmid purification

After chemical transformation, single colonies were picked and added to the LB medium or 2YT medium with the appropriate ampicillin amount. The cells were incubated overnight at 37 °C at 130 – 200 rpm. The plasmid purification was done using Monarch® Plasmid Miniprep Kit (New England Biolabs). 2 – 4 ml of the bacterial cell was centrifuged for 30 sec then the supernatant was discarded. Pellet was resuspended in 200 µl resuspension buffer then mixed, 200 µl of lysis buffer was added then genteelly mixed, 400 µl of neutralization buffer was added then incubated for 2 min. The lysate was clarified by spinning at 16,000 x g for 2 – 5 min. The supernatant was transferred to a spin column and centrifuged for 1 min, then 200 µl of primary wash buffer was added and then centrifuged for at 16,000 x g 1 min. Afterwards, 400 µl of secondary wash buffer was added then centrifuged again at 16,000 x g 1 min. Finally, 20 – 30 µl of elution buffer was added to elute the DNA.

### 2.2.1.7. Restriction digest of DNA

Digestion of purified DNA or plasmid DNA of PCR product was used for analysis, checking and/or cloning by using NEB FastDigest enzymes (New England Biolabs) or FastDigest enzymes (Thermo Fischer Scientific). For a preparative digest reaction, 2 µl of 10x FastDigest buffer, 1 µl of each of the two FastDigest restriction enzymes, <1µg of DNA and 20 µl water was added. For analytical test digestion, 1 µl of 10x FastDigest buffer, 0.3 µl for the FastDigest enzymes, 2 µl of DNA and 6.5 µl water was added.

### 2.2.1.8. Polymerase chain reaction

For linearization, amplification and isolation of DNA fragments, Polymerase chain reaction (PCR) was used. Typical PCR was done using the Q5 high fidelity DNA polymerase or Taq DNA polymerase (New England Biolabs) after designing primer pairs that target a specific DNA fragment using Snapgene (GSL Biotech LLC). All the primers were ordered from Sigma/Merck (see 2.1.7.1.). The colony PCR was done by picking single colonies with a pipet tip which is transferred into the reaction mix. The component of a typical PCR

component is shown in table 17, while the program used for the PCR cycles is shown in table 18.

**Table 17. Master mix reaction for PCR.** Tables include components of Q5 DNA polymerase and Taq DNA polymerase

| Component | Q5 DNA polymerase |
|---|---|
| 5x Q5 reaction buffer | 4 µl |
| 10 µm forward primer | 1 µl |
| 10 µm reverse primer | 1 µl |
| 10 mM dNTPs | 0.4 µl |
| Template DNA | 1 pg – 1 µg |
| Polymerase used | 0.2 µl |
| Nuclease-free water | up to 20 µl |

| Component | Taq DNA polymerase |
|---|---|
| 10x Taq reaction buffer | 2 µl |
| 10 µm forward primer | 1 µl |
| 10 µm reverse primer | 1 µl |
| 10 mM dNTPs | 0.4 µl |
| Template DNA | 1 pg – 1 µg |
| Polymerase used | 0.1 µl |
| Nuclease-free water | up to 20 µl |

**Table 18. PCR protocol for Q5 DNA polymerase and Taq DNA polymerase.** The table includes the temperature, time and cycle of each step

| Q5 DNA polymerase | | | |
|---|---|---|---|
| | **Temperature** | **Time** | **Cycle** |
| **Step** | | | |
| **Initial step** | 98 °C | 30 sec | |
| **Denaturation** | 98 °C | 5 – 10 sec | Repeat for 20 |

| Taq DNA polymerase | | | |
|---|---|---|---|
| | **Temperature** | **Time** | **Cycle** |
| **Step** | | | |
| **Initial step** | 95 °C | 30 sec | |
| **Denaturation** | 95 °C | 5 – 30 sec | Repeat for 20 |

| | | | |
|---|---|---|---|
| **Annealing** | 50 – 70 °C | 10 – 30 sec | – 35 cylces |
| **Extension** | 70 °C | 30 sec per 1 kb | |
| **Final extension** | 70 °C | 2 min | |
| **Hold** | 4 – 12 °C | ∞ | |

| | | | |
|---|---|---|---|
| **Annealing** | 50 – 70 °C | 5 – 60 sec | – 35 cylces |
| **Extension** | 70 °C | 60 sec per 1 kb | |
| **Final extension** | 70 °C | 5 min | |
| **Hold** | 4 – 12 °C | ∞ | |

## 2.2.1.9. Seamless cloning and DNA assembly

Cloning and DNA assembly were carried out using the NEBuilder HiFi Assembly Master Mix (New England Biolabs), which is also known as Gibson assembly, by designing primers from Snapgene (GSL Biotech LLC). The primers were designed to have an overall $T_m$ of 50 - 60 °C depending on the overlapping bases. The reaction was done following the manufacturer's recommendation of the molar ration, vector: insert ratio of 1:2, and the total amount of fragments of 0.03 to 0.2 pmols. Afterwards, the sample was incubated for one hr at 50 °C. Additionally, NEBuilder (New England Biolabs) was used for mutagenesis by creating primers with the desired mutation. All transformations of final cloning steps were carried out as described in section 2.2.1.1.

## 2.2.1.10. Site-directed mutagenesis

A Q5 Site-Directed Mutagenesis kit (New England Biolabs) was used for single or double amino acid residue exchange. The PCR mix and cycles are pointed out in section 2.2.1.8. After the PCR, 1 µl of KLD enzyme mix (New England Biolabs) is composed of Kinase, Ligase and DpnI. The reaction was followed up by a transformation which is described in section 2.2.1.1. Primers were created by either NEBaseChanger (New England Biolabs) or Snapgene (GSL Biotech LLC).

### 2.2.1.11. DNA gel extraction

Gel extraction of DNA was done using Monarch DNA gel Extraction Kit (New England Biolabs). DNA extraction was done by following the manufacturer's protocol. Samples were incubated at 50 – 55 °C prior to DNA elution. DNA was eluted with 8 to 25 µl of elution buffer.

### 2.2.1.12. DNA sequencing and analysis

In order to validate the DNA sequences, Sanger sequencing was done by Microsynth SeqLab and Eurofins Genomics companies. The sequencing was done using custom-designed primers or slandered primers offered by the companies. Custom made primers were designed either manually while checking $T_m$ online with NEB $T_m$ calculator or via Snapgene (GSL Biotech LLC). The sequences were analyzed using Chromas (Technelysium), Multialign (Florence Corpet), or Snapgene(GSL Biotech LLC).

## 2.2.2. Protein biochemical methods

### 2.2.2.1. Protein expression

For the general recombinant protein expression, recombinant genes were expressed in the different strains shown in 2.1.10. Cells were picked directly from already prepared glycerol stocks or freshly transformed cells on LB-Amp plates. Primarily, 400 ml of liquid 2YT or LB was autoclaved and pre-warmed before usage with the appropriate antibiotic concentration of 50 - 100 mg/ml. Pre-cultures were prepared by inoculating single colonies to 60 - 40 ml of sterile LB or 2YT that was previously prepared. The optical density was recorded, and the appropriate amount of pre-culture was used based on the optical density (OD). The amount was calculated by adjusting the appropriate volume, in millilitres, for an $OD_{600}$ of 2.0 as a standard. For example, 5 ml of pre-culture at $OD_{600}$ 2.0 means 5.263 ml for pre-culture at $OD_{600}$ of 1.9. To start the expression, 500 µl of ampicillin was added, of a stock of 50 mg/ml, in pre-warmed 500 ml autoclaved 2YT or LB medium. An appropriate volume of pre-culture

added, then the flasks were incubated in Multitron standard (INFROS HT) at 28 to 37 °C at 130 rpm (see table 16). At $OD_{600}$ of 0.5 to 0.6, 1mM final concentration of sodium molybdate or $Na_2MoO_4$ (Sigma-Aldrich) was added. This was done using 500 µl of a 1M stock solution. Afterwards, the temperature was lowered to 22 °C until reaching $OD_{600}$ of 0.8 – 1.0, was cells was induced by adding 1 mM final concentration of IPTG (500 µl of 1M stock solution or 5 ml of a 100 mM stock solution). The expression cultures were incubated overnight at 22 °C at 130 rpm.

Recombinant expression of both Cnx1 and gephyrin full-length was done in the *E. coli* strain BL21 DE3 or TP1000 (see 2.1.10.). The initial recombinant expression procedure was the same as the general protein expression. However, the temperature was lowered to 19 °C in case of Cnx1 full length and 16 °C in case of gephyrin constructs and kept overnight in that temperature after induction (0.5 mM and 1 mM final concentration of IPTG for Cnx1 full length and gephyrin full-length, respectively). Additionally, no sodium molybdate was added for gephyrin. However, 1 mM final concentration was added to Cnx1. All $OD_{600}$ was recorded using a UV/Vis Spectrophotometer Ultrospec™ 2100 pro (Amersham Bioscience) or Lambda 25 (PerkinElmer) which was pre-equilibrated and distilled water used as a blank for the machine. Afterwards, cells were harvested by centrifugation in Heraeus™ Multifuge™ X3R (Thermo Fischer Scientific) for 20 min at 4,500 xg or Avanti J-25 (Beckmann) for 8 min at 8,000 xg with a temperature generally set at 4 °C. The cell pellet was resuspended in the appropriate pre-cooled lysis buffer depending on the subsequent step (5 – 3 ml for each 1g of cell pellet). Subsequently, cells were either shock frozen in liquid nitrogen and stored at -80 °C or stored directly at -20 °C.

## 2.2.2.2. Protein purification

### 2.2.2.2.1. Cells Lysis

For cell lysis, initially, cells thawed at room temperature or with a Mini Therm1 water bath (DINKELBERG). Afterwards, a tip of a spatula of DNase I (Roche) was added, and the cells was incubated for 5 min. In the case of full-length proteins (Cnx1 or gephyrin), one cOMPLETE tablet (Roche) was crushed and added for every 50 ml of resuspended. Lysis was

performed twice by either French® press cell (SLM Instruments) or EmulsiFlex-C5 Homogenizer (Avestin) at 1,000 – 1,500 psi. Next, lysed cells spanned down at 22,222 xg for 45 - 60 min at 4 °C, and the supernatant was used for further affinity chromatography purification step (See 2.2.2.2.2.). For small volumes of E. coli or *Neurospora crassa,* FastPrep-24™ was used (MP Biomedicals) twice for 30s at 6,5 m/s.

## 2.2.2.2.2. Affinity chromatography

For Strep-tag® purification, recombinant genes were fused with Strep-tag® or Twin-strep sequence at the C- or N- terminus of the gene. All affinity chromatography steps were performed at 4 to 6 °C. Firstly, 1 or 2 ml of Strep-tactin® (IBA) resin with lysis buffer (see 2.1.6.) in a 10 or 25 ml column (Bio-Rad) was prepared for purification. The resin was equilibrated 10 times each 1 ml of resin with the appropriate pre-cooled lysis buffer. The supernatant was then loaded on the resin, and the resin was washed 10 times per 1 ml of resin with the pre-cooled wash buffer (e.g. 1 ml resin was equilibrated with 10 ml lysis buffer and washed with 10 ml of wash buffer) (see 2.1.6.). The protein elution was done using 6 – 10 ml of 5 or 25 mM D-desthiobiotin (IBA) dissolved in the wash buffer. The eluted proteins were concentrated using Vivaspin® Turbo 4 (Sartorius) spin columns that were pre-equilibrated with wash buffer. The proteins were concentrated to a volume of 100 to 500 µl. Proteins were then shock frozen in liquid nitrogen and stored in LS 750 liquid nitrogen tank (TechLab Gmbh) or used directly for further protein purification steps (see 2.2.2.2.4. and 2.2.2.2.5.) or cross-linking (see 2.2.4.).

## 2.2.2.2.3. Protein concentration determination

For protein concentration determination of purified proteins, $A_{280,}$ or $OD_{280}$, was used. Proteins were added on a plate reader that was inserted in Multiskan™ GO plate reader (Thermo Fisher Scientific), and the extinction coefficient was obtained by submitting the sequence on ExPASy ProtParam (https://web.expasy.org/protparam/). The protein concertation was finally calculated using the following formula C = A280 /ε b * 20 * MW (C for concentration, ε for extension coefficient, b is path length, and MW is the molecular

weight). On the other hand, for a mixture of protein or crude extract of *E.coli* and *N. crassa*, Bradford protein assay was used. For Bradford protein assay, different dilution of the protein solution was added to a 1 ml of a Roti Quant (Carl Roth) which was diluted from a 5x stock solution. The protein samples were incubated at room temperature for 10 – 15 min, and the $OD_{595}$ was measured using the UV/Vis spectrophotometer Ultrospec™ 2100 pro (Amersham Bioscience).

## 2.2.2.2.4. Size exclusion chromatography

Further purification of proteins/ protein complexes was carried out using gel filtration or size exclusion chromatography (SEC) were proteins are separated based on size or hydrodynamic radius. In SEC proteins with large molecular weight elute faster than smaller ones (Brusotti *et al.,* 2018). For this purpose, firstly, the appropriate column (shown in table 19) were connected to the ÄKTA basic or ÄKTA purifier system (Amersham Pharmacia Biotech). Secondly, the column was equilibrated with the appropriate buffer with a flow rate shown in table 19. Buffers used were shown in 2.1.6., and all the buffers used were filtered and degassed for SEC. Fractions of the protein of interest were taken, and the fraction was analyzed using SDS PAGE (see 2.2.2.3.1), or the proteins were concentrated for further uses by using the Vivaspin® 15 10,000 MWCO (Sartorius). Proteins were then stored on ice for further experiments or shock frozen in liquid nitrogen and stored in the protein storage tank.

**Table 19. Columns used for gel filtration.** The table includes the name of the column, pressure limit, bed volume, bed dimension, flow rate used, sample volume injected and purpose of use

| | **Superdex 200 Increase 10/300 GL** | **Superdex 200 10/300 GL** | **Superdex 75 Increase 5/150 GL** |
|---|---|---|---|
| **Pressure limit** | 3 MPa | 1.5 MPa | 3 MPa |
| **Bed volume** | Approximately 24 ml | Approximately 24 ml | Approximately 3 ml |
| **Bed dimension** | 10 x 300-310 mm | 10 x 300-310 mm | 5 x 150 mm |
| **Flow rate used** | 0.3 – 0.5 ml/min | 0.3 – 0.5 ml/min | 0.2 – 0.4 ml/min |
| **Sample volume injected** | 100 – 500 µl | 100 – 500 µl | 10 – 50 µl |
| **Purpose** | Preparative uses or to purify proteins | Preparative uses or to purify proteins | Analytical uses or to visualize, analyze |

| | | | |
|---|---|---|---|
| | | | purity and estimate native protein sizes |

## 2.2.2.2.5. Ion-exchange chromatography

Further polishing and separating of proteins/ protein complexes were achieved via Ion exchange chromatography (IEX). For this work, Anion exchange was performed using SUORCE™ 15Q (GE Healthcare Life Sciences) using a pH that was calculated based on the isoelectric point (pI) of the protein. The pI of the proteins was obtained by submitting the amino acid sequence to ExPASy ProtParam (SIB). The buffer was prepared with a pH of at least 1 pH unit (e.g. for a pI of 5.5, 7.2 pH was used) (see 2.1.6.). The proteins were eluted by increasing salt concentration, from 0 - 10 mM salt to 1M salt gradient, for 60 to 80 min. Before running IEX, PD10 columns (GE Healthcare Life Sciences) was used to buffer exchange to the buffer with the appropriate pH. In the case of an IEX following a SEC step, the low/no salt buffer was used already in SEC, and there was no need to do another buffer exchange step using PD10 columns. More detailed information about SEC is included in the previous section (see 2.2.2.2.4.). All buffers were prepared as described in section 2.1.6.

**Table 20. Anion exchange chromatography parameters.** The table includes information on the ion-exchange runs done

| Anion exchange chromatography parameters | |
|---|---|
| **Column used** | SOURCE™ 15Q |
| **Starting salt concentration** | 0 – 10 mM |
| **Ending salt concentration** | 1M |
| **Injected volume** | 2 to 3.5 ml |
| **Equilibration flow rate** | 2-3 ml/min |
| **Post- sample injection flow rate** | 1 ml/min |
| **Gradient time** | 60 to 80 min |
| **Starting percentage of first buffer (A)** | 0 to 10% |
| **Ending percentage of second buffer (B)** | 100% |
| **pH used** | 7.2 |
| **Fractions volume taken** | 500 µl |

| | |
|---|---|
| **Storage condition** | 20% Ethanol |

## 2.2.2.3.Protein characterization

### 2.2.2.3.1. SDS-PAGE

In order to separate, analyze and evaluate purified proteins or cell lysate, sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) was routinely used. Samples loaded on the acrylamide gels were separated based on their electrophoretic mobility. For this purpose, samples were prepared in 2.5 or 5x loading buffer (see 2.1.6.), and a protein standard was loaded as a reference (see 2.1.5.). For cell lysate and crude extract, samples were heated up to 90 °C for 10 min before loading on the gel. However, for proteins, boiling or heating up was not done to avoid aggregation of the protein complexes/ samples. In order to run the SDS gels, two different systems were used to run SDS-PAGE. For the regular SDS gels using Julaloo F10 (Julaloo) with PHERO-stab.300 power supply (BIOTEC fischer) and polyacrylamide gels of 7.5 - 15 % (w/v) were prepared. The gel electrophoresis ran at 200 V for 45 - 90 min, and then the gels were stained overnight with Serva Blue R (SERVA) or Serva Blue G (SERVA). The next day, the gels were destained using either 40% ethanol or destaining solution for 1 hr. Ponceau S was used for extended storage purposes to stain the gels after being transferred on a polyvinylidene fluoride (PVDF) membrane (see 2.2.2.3.2). On the other hand, for a quick staining-free gel, TGX Stain-Free™ FastCast™ Acrylamide Solutions (BioRad) following the manufacturer's protocol. Stain-free gels ran using Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (BioRad) with PowerPac HC (BioRad) power supply at 250 V for 30 - 40 min. In both systems, gel visualization and documentation were done with ChemiDocTM XRS+ (Bio-Rad Laboratories) imager.

### 2.2.2.3.2. Western blot

For Western blot analysis, the SDS gels prepared (see 2.2.2.3.1) was not stained. The gels were transferred directly on pre-equilibrated PVDF membranes in transfer buffer (see 2.1.6.) after the gel electrophoresis. For semi-dry Western blotting, Trans-Blot Turbo Transfer System

(Bio-Rad Laboratories) was used according to the manufacturer's guidelines. Afterwards, the membrane was blocked in 5% (w/v) skimmed milk (Heirler) powder in TBS-T buffer (see..) for 1 hr on the W55 (Edmund Bühler) shaker at room temperature or in the cold overnight on 3015 GFL (Wilh. O. Schmidt Gmbh) shaker. The membrane was washed 3 times with cold TBS-T for 5 min while shaking then the appropriate primary antibody was added (see 2.1.9.) for 1 hr. After incubation, the primary antibody was collected, and the membrane was washed 3 times with cold TBS-T for 5 min while shaking. Subsequently, the membrane was incubated with 20 ml of 5% (w/v) skimmed milk for 10 min, and then 2 µl (1: 10,000 dilution) of the (peroxidase-conjugated AffiniPure Goat Anti-Mouse IgG + IgM (H+L), Dianova) secondary antibody was added and incubated for 1 hr shaking at room temperature. The membrane was washed 3 times with TBS-T, and proteins were visualized and detected using SERVALight Eos (SERVA) and ChemiDocTM XRS+ (Bio-Rad) to image the membranes.

#### 2.2.2.3.3. Native PAGE

Typically, the proteins were mixed together for few hrs or overnight prior to the run. Native gels were casted using native gel buffer and submerged in 12.5 mM Tris pH 8-9, 96 mM glycine buffer. The whole gel cassette was placed in an icebox with an active cooling system (Julaloo). The gels were equilibrated for 30 min prior to the run. The gel electrophoresis was done using Julaloo F10 (Julaloo) with PHERO-stab. 300 power supply (BIOTEC fischer) at 50 V for 3 to 4 hrs then the gels were stained overnight with Serva Blue R (SERVA) or Serva Blue G (SERVA).

### 2.2.3. Quantification of MPT, MPT-AMP, Moco, and Moco-AMP metabolites

For the quantification of MPT-AMP or Moco-AMP and Moco content, as in an already published protocol (Krausze *et al.,* 2017) (Figure 10), proteins were initially set to a 10 mg/ml concentration (using A280 – see 2.2.2.2.3.). Subsequently, 2, 5 and 10 µl of the 10 mg/ml concentrated proteins were added to an oxidation solution (800 ml of 0.1 M Tris/HCl (pH

7.2) + 50 ml of acidic iodine (1% (w/v) $I_2$/2% (w/v) KI in 1 M HCl) which convert the Moco content into Form A. The Form A sample was then divided into two, 450 µl of the sample with only Alkaline phosphatase to remove the phosphate group and another 450 µl for Phosphodiesterase and alkaline phosphatase to remove both phosphate group and the phosphodiester bond. Form A samples ran on an HPLC Agilent 1100 using ReproSil-Pur Basic-C18-HD (TECHLAB Gmbh) with 250 x 4.6 mm dimensions and 5 µm bead size, as well as a fluorescence detector. A synthetic Form A standard was used as a reference.

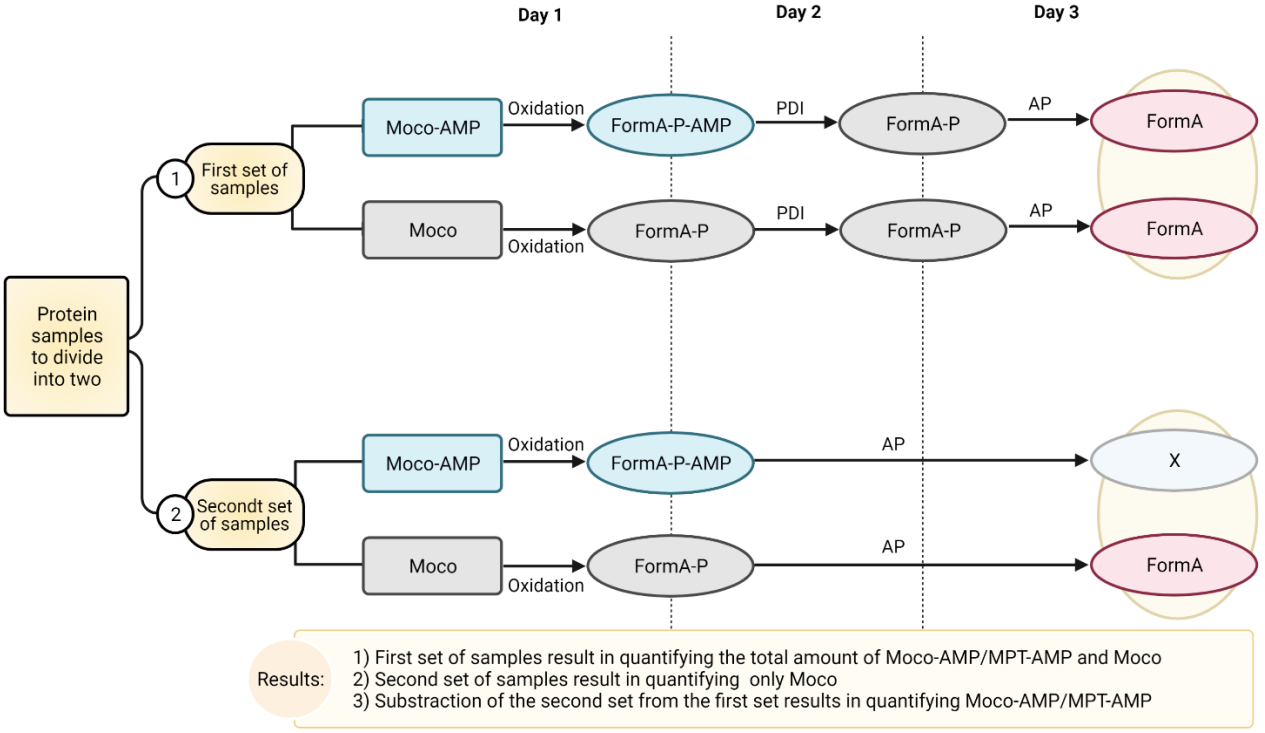



**Figure 10. FormA based quantification of MPT, MPT/Moco-AMP and Moco.** The protocol starts on day 1 with the splitting of the samples into two sets. Both sets are oxidized by adding acidified $I_2$/KI solution to the samples. For Moco-AMP or MPT-AMP, oxidation will yield FormA with phosphate and AMP bound. For Moco, oxidation will leave it with the only phosphate. On Day 2, phosphodiesterase I (PDI) is added for the first set, which will remove AMP yielding phosphorylated FormA (FormA-Phospho). For the second set, alkaline phosphatase (AP) is added, which results in the removal of the terminal phosphate group from FormA but not FormA-AMP. On Day 3, AP is added to the first set yielding FormA dephospho for the total quantity of metabolites. After quantification using the HPLC (only detects formA), subtracting the results of the second set (only the Moco content but not the adenylated metabolite) from the total metabolite quantity resulting from the first set (Both Moco and Moco-AMP/ MPT-AMP), quantifying Moco-AMP/ MPT-AMP is possible (Hercher *et al.,* 2020).

## 2.2.4. Chemical crosslinking

Chemical crosslinking of proteins is done in order to stabilize protein complexes, transient or non-transient complexes, to take a ''screen shot'' of how complexes can possibly form.

This facilitates our understanding about such complexes by determining their structure (Sinz, 2006). The general *in-vitro* protocol for cross-linking with DSS, DFDNB, DST, DSSO and DSBU is, initially, a 10 min equilibration of the cross-linker at room temperature (see table 21). This is due to the moisture sensitivity of such chemicals and to prevent moisture condensation in the bottle or vial. A stock solution of 50 mM or 1M solution of the cross-linker is prepared or approximately 10 mg of the cross-linker in 500 µl DMSO (depending on the molecular weight of the cross-linker). The molar excess of the cross-linker is of absolute importance. The minimal molar excess used in this work is 10-fold excess (1:10 protein to cross-linker). The protein is added to the appropriate cross-linker in HEPES buffer (a non-amine containing buffer) at an optimal pH of 7.2 (7 – 9 pH range) (see 2.1.6.). The reaction was left at room temperature for 25 – 30 min or 2 hrs on ice. The reaction was quenched afterwards by adding 50 to 200 mM Tris-HCl and incubated for 10 – 15 min. Another approach to eliminate excess cross-linker is either by using PD10 columns (for further IEX purification) or loading the proteins on the SEC for downstream procedures. For *in-vivo* cross-linking, the same procedure as the *in-vitro* cross-linker, however, the crosslinking was done to *N. crassa* or *E.coli* before lysis. After the crosslinking reaction is quenched, cells were lysed using FRENCH® pressure cell press (SLM Aminco®) or FastPrep-24™ was used (MP Biomedicals) twice for the 30s at 6,5 m/s (see 2.2.2.2.1.).

**Table 21. Cross-linkers used in this work.** The table contains information about the type and purpose of the cross-linkers used, including properties and subsequent techniques used

| | **Crosslinking conditions for visualizing complexes, initial tests and optimization for other techniques** | **Crosslinking for Mass Spectrometry** | **Crosslinking for Negative staining and Cryo-Electron microscope** |
|---|---|---|---|
| **Cross-linker(s) commercial name(s)** | 1. Formaldehyde<br>2. DSS<br>3. DFDNB<br>4. DST<br>5. EDC<br>6. DSSO<br>7. DSBU | - DSBU | - DSS |

| | | | |
|---|---|---|---|
| **Bi-functional properties** | Amine to amine crosslinking except for EDC, which has carboxyl- to an amine group | Amine to amine crosslinking | Amine to amine crosslinking |
| **Subsequent technique(s) used** | SDS PAGE and SEC | SDS PAGE and LC-MS/MS | SDS PAGE, SEC, IEX and Cryo-electron microscope |
| **Purpose of use** | Analyze and visualize stabilized protein complexes/ cross-linked products | Determine amino acids that are in close-proximity in space but not in sequence, which leads to the development of a low-resolution model | Obtain/ determine protein complex structure |

After crosslinking this, the cross-linked proteins were loaded onto a 4-20% polyacrylamide gradient gel (Sigma Aldrich) or TGX-stained gels (see 2.2.2.3.1). The protein complexes' bands were excised from the gel, then reduced with excessive dithiothreitol (DTT) and carbamidomethylated with iodoacetamide (IAA). Subsequently, the samples were incubated overnight with GluC (1:20, enzyme: protein ratio) at 37 °C. Subsequently, a 4 hr trypsinization step was carried out (Promega).

## 2.2.5. Mass spectrometry

## 2.2.5.1. LC/MS/MS

For LC/MS/MS, a C8 reverse phase (RP) (Thermo Fisher Scientific) was used as a precolumn. C18 RP (Thermo Fisher Scientific) was used as a separation column. For the nano-HPLC, the HPLC was coupled to the nano-ESI source of the Orbitrap Fusion (Thermo Fisher Scientific). To desalt the samples, 0.1% trifluoroacetic acid (TFA) was used on the pre-column for 15 min. LC-MS grade H2O (VWR) with formic acid (FA) was used as solvent A, and LC-MS

grade of 80% acetonitrile (VWR) and 0.08% FA was used as solvent B. An elution gradient of flow rate of 300 ml/min was set to 35% of solvent B in 90 min. For data acquisition, data-dependent MS/MS mode was applied, and Xcalibur (Thermo Fisher Scientific) was used to control the data acquisition.

### 2.2.5.2. Analysis of Results

Cross-linked products were identified using MeroX (version 1.6.6) (Iacobucci *et al.,* 2018). Data analysis was done using Mascot generic format (mgf) files. To define the enzymatic cleavage sites, *C*-terminal to K and R was set for trypsin and *C*-terminal to D and E was set for GluC. For amino acid residues K, R, D and E, three missing cleavage sites were allowed. The cross-linker name was identified as DSBU with composition C903NH12, and then the cross-linking site one was defined for K and site two was defined for K, S, T, Y and *N*-terminal. Consecutive peptides as cross-links were ignored. For mass comparison, the lower mass limit was set to 200.0 Da and the upper mass limit of 6000.0 Da with ion types a, b and y. Rise mode was set on, as well as the cRAP database included. All cross-links generated by the software were manually inspected and visualized on the 3D protein structure by PyMOL (Schrödinger LLC).

### 2.2.5.3. Native MS

The proteins were incubated together in equimolar ratio for 60 min then the buffer was exchanged to 500 mM ammonium acetate, pH 6.8, with Amicon Ultra centrifugal unit of 10 kDa cutoff (0.5 mL) before the measurements of the native MS. Using a modified high-mass QToF II instrument (Waters Micromass/MS Vision), Native MS measurements were obtained. Cone voltage was set to 160 V, source voltage to 1400 V and extraction cone voltage to O V (various collision energies of 30 and 50 V was done).

## 2.2.6. Crystallization, data collection and model building

Before crystallization, MogA and MoeA were concentrated to about 40 mg/ml using Vivaspin concentrator columns with a 10 kDa molecular weight cut-off. 0.015 M of $MgCl_2$ and adenosine diphosphate (ADP). MoeA and MogA crystals were obtained from various screen conditions. Protein crystals were flash-frozen in liquid nitrogen prior to X-ray diffraction experiments
on beamline P11, operated by DESY at the PETRA III synchrotron (Hamburg, Germany). autoPROC was used to process the data set (Vonrhein *et al.,* 2011), and the anisotropy was corrected with STARANISO (Tickle *et al.,* 2017), while the crystallographic phase problem was fixed with Phaser (McCoy *et al.,* 2007). Structure refinement was implemented using BUSTER (Bricogne *et al.,* 2017) and the structure was rebuilt using Coot (Emsley *et al.,* 2009).

## 2.2.7. Computational Modelling

## 2.2.7.1. Homology modelling and *Ab-initio* modelling

Computational docking methods have shown to be helpful in structural prediction. While homology modelling is template-based modelling or local docking, *ab initio* modelling is constructing models from scratch (Marsh and Teichmann, 2015). For homology modelling, Phyre2 (Kelley *et al.,* 2016) was used. Phyre2 is an automated fold-recognition server for the prediction of the structure of proteins. Initially, PDB files were edited by Coot (Emsley *et al.,* 2009), if necessary, and uploaded on the server for intensive prediction protocol. Afterwards, the generated models were checked using FiberDock (Mashiach *et al.,* 2010). FiberDock is a method for the flexible refinement of docking proteins. All homology and Ab-initio models were uploaded to FiberDock for another final flexibility refinement. For *Ab-initio* modelling, I-TASSER (Roy *et al.,* 2010) or Robetta (Simons *et al.*, 1999) was used. Sequences were uploaded to I-TASSER or Robetta, subsequently, a pre-edited PDB file with PyMOL (Schrödinger, Inc.) is uploaded for the known structure of the sequence. The result is 5 ranked models which are not template-based but *de novo*.

### 2.2.7.2. Global docking

For global docking (no restraint protein-protein docking), ClusPro was used (Kozakov *et al.,* 2017). ClusPro relies on an unprejudiced selection of protein docking without any previous information provided about the individual proteins. Firstly, the PDB structures were prepared and edited if necessary, using Coot (Emsley *et al.,* 2009). Afterwards, the first protein was added as a receptor, and the second was used as a ligand (or vice versa). The modelling is done on three different stages: 1) Rigid body stage where billions of conformations are used for sampling, 2) Root-mean-square deviation (RMSD) using clustering of lowest thousand energy structures, 3) final refinement of selected structures and subsequent energy minimization. In the first stage, PIPER is used as a docking program relying on Fast Fourier Transform (FFT) (Kozakov *et al.,* 2017). Initially, the first protein, MoeA in this case, is used a receptor in which the protein is placed on a fixed grid at the origin of the coordinate system. The second protein, MogA, is placed on a movable grid as a ligand. The correlation function is how the interaction energy is written, and the energy functions can be calculated precisely using FFT, which results in billions of conformations of the two interacting proteins. This enables the unbiased blind selection of protein docking without any previous information about the individual proteins. Secondly, the 1000 models with the lowest energy based on the root-mean-square deviation (RMSD) clustering are selected. The structures are calculated based on finding the structure within 9Å RMSD with the highest number of neighbours. The neighbours will then form the second cluster, and up to 30 clusters are generated in the same manner. Lastly, using energy minimization, structures are refined, and the energy is minimized for the retained structures for 300 steps, resulting in few conformational changes (Desta *et al.,* 2020). The output of ClusPro is the 10 most populated clusters selected based on electrostatic-favored, hydrophobic-favored, and van der Waals electrostatics. Models were all manually checked using PyMOL (Schrödinger, Inc.) or Chimera (RBVI).

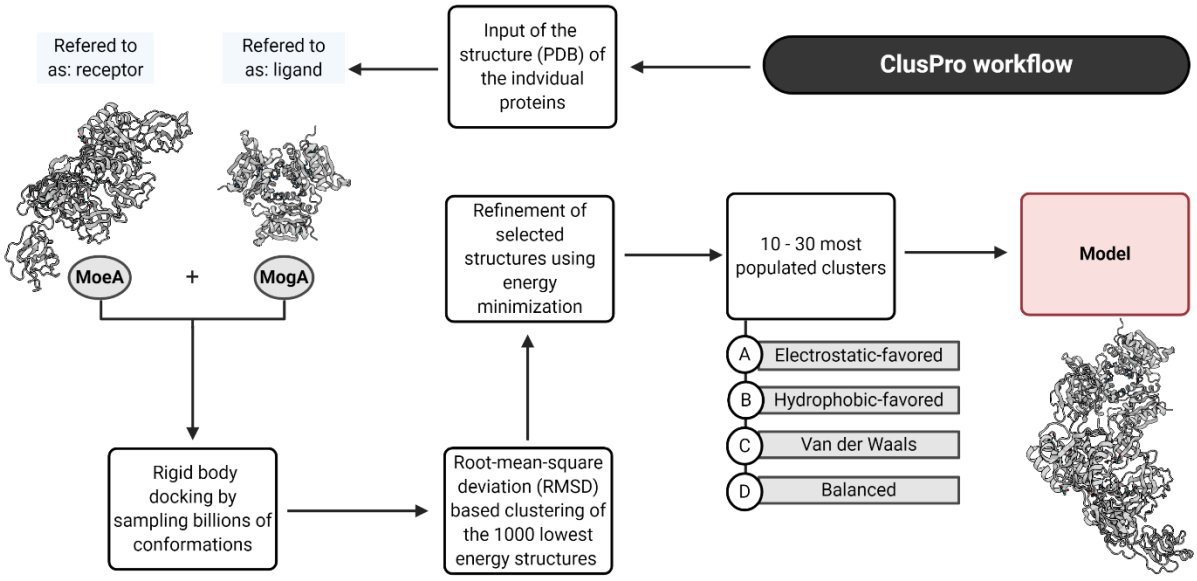



**Figure 11. ClusPro Workflow.** Brief summary of the workflow of ClusPro, starting with the input of structures to generation of the models.

Finally, the models obtained from Cluspro were constantly refined using FiberDock (Mashiach *et al.,* 2010) for flexible refinement of docking proteins to consider the protein flexibility. FiberDock relies on modelling both side-chain and backbone flexibility simultaneously as well as performing rigid body optimization on the orientation of the ligands. This is important mainly for the flexible parts of the protein structure in which they undergo local and global conformation changes, including opening binding sites and movement of loops. A side-chain rearrangement subsequently refines the structures in a restricted interface manner (Mashiach *et al.,* 2010). FiberDock was used in the case of ClusPro protein-protein docking since it adds some flexibility to the protein docking. However, this was not necessarily required for HADDOCK modelling since HADDOCK automatically has this integrated into one of the modelling stages.

### 2.2.7.3. Modelling the Cross-links

Modelling with restraints can be done by many different modelling software. In this work, high ambiguity driven protein-protein docking (HADDOCK) was used (Koukos et al., 2020). A summary of the crosslinking-based modelling workflow is shown in figure 12. Initially, PDB files are prepared by removing all water molecules and ligands using PyMOL. Further structural editing was done using Coot (Emsley et al., 2009) in order to renumber sequence or chains as well as the identifier. Using PHENIX (Liebschner et al., 2019), all alternative conformations were removed, which is necessary to be done or else HADDOCK will not accept the structure input files. Afterwards, TBL files are prepared by arranging all the restraints in order after manually checking them using PyMOL (Schrцdinger, Inc.). TBL files are files containing information that the server needs to start the modelling. The information includes the residues numbers, locations and specific distance range. The restraint distance of Cα to Cα is added in a range of 4 to 30 E distance (DSBU cross-linker span). It is crucial to prepare the TBL files correctly, or else the modelling will fail. For instance, the numbering of the PDB sequences, which is slightly different from the actual protein sequence, should always be manually checked prior to the preparation of the TBL files. Cport (Vries et al., 2011), which contains six prediction methods, is used to identify all active and passive residues of the individual PDB files. Once the PDB and TBL files are uploaded as well as inputting all active and passive residues, the modelling can start. HADDOCK modelling goes through 3 stages, firstly, energy minimization of rigid-body, secondly, semi-flexible simulated annealing and finally, flexible refinement in a water shell. This result in different clusters which all have different Z-scores and RMSD. The lowest Z-score model is used, which is checked manually using PyMOL. In case of any unreliable or odd results, TBL files were re-checked and edited before re-modelling again.

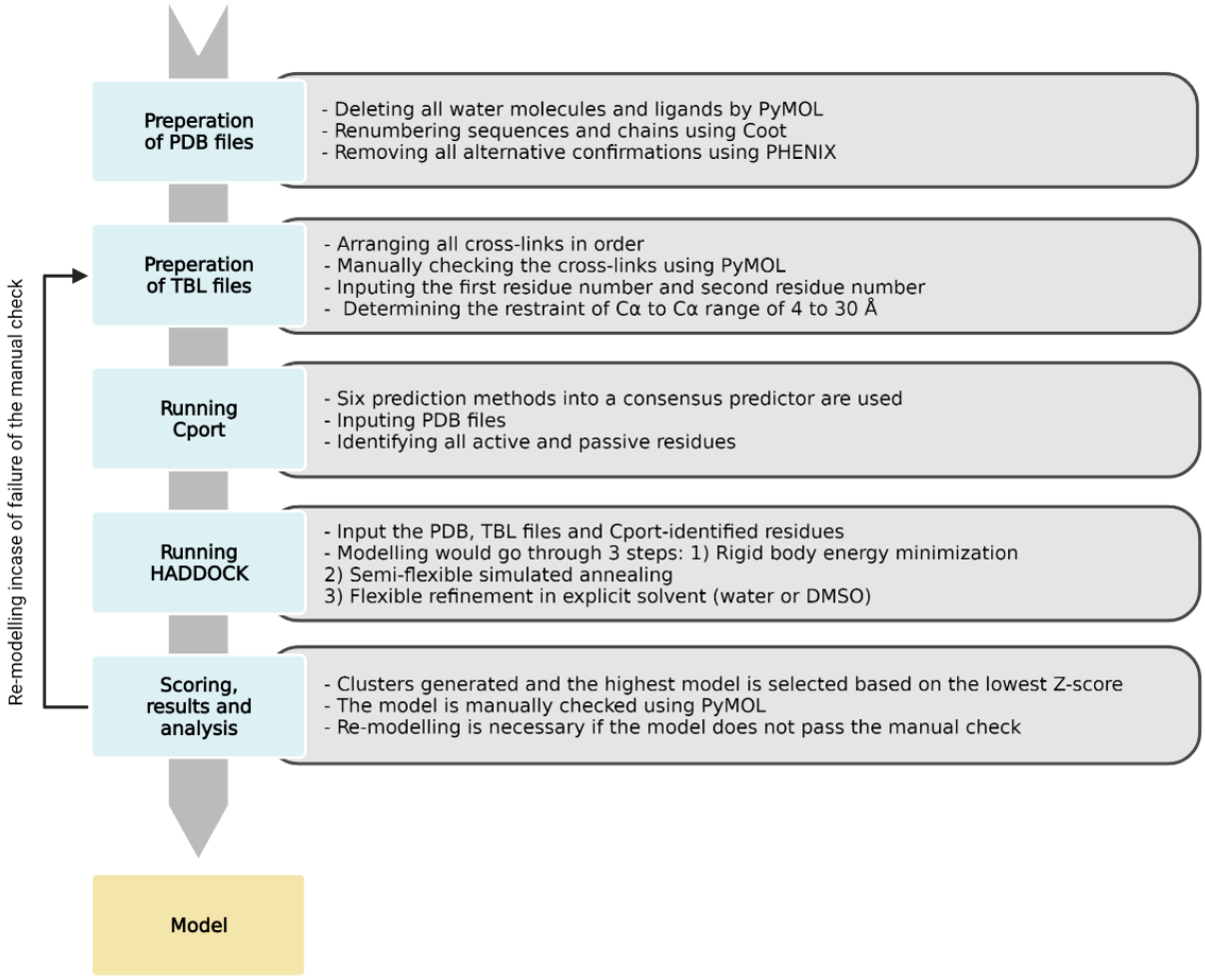



**Figure 12. Work flow of modelling restraints with HADDOCK**. The work flow starts with preparing the PDB files, preparing the TBL files, running Cport, running the modeler, analysing, and generating models. The models are subjected to manual investigation, resulting in re-modelling after a manual quality control check.

## 2.2.8. Electron Microscopy and negative staining

Negative staining is a technique that is often used in microscopy to visualize and analyze specimens. The samples are stained with a heavy metal buffer, such as uranyl acetate, which stains all the background while leaving the actual samples unstained and visible (Ohi *et al., 2004*). Negative staining was done at ≈ 20nM proteins in 150 mM HEPES, 300 NaCl buffer. The proteins were centrifuged for 20 min at 14,000 gx before negative staining. Carbonfilm on 400-square-mesh copper grids were plasma-cleaned using a plasma cleaner (Diener)

which was then stained with 2% uranyl acetate (Sigma) twice with a 15-second pause in-between and dried with filter paper. Using TALOS FC200 (Thermo Fisher Scientific), ≈ 200 electron micrographs were recorded. EPU thermos, ATLAS and TEM user software (Thermo Fisher Scientific) were used to take the micrographs and navigate the microscope. ≈ 10,000 partials were picked manually using RELION (Scheres 2012). 2D and 3D classification was done, which then was visualized using Chimera (RBVI). The crystal structures were fitted in the models using Chimera, and all fitting was done automatically by fitting the PDB structures to the EM map using the match and fit to map command.

# Chapter three

# 3. Results and Discussion

## 3.1. MoeA-MogA complex

### 3.1.1. Overview

MoeA and MogA complex is considered to be the early form of the evolutionary Mo-insertase complex prior to the introduction of the linker between both proteins (Krausze *et al.,* 2018). Therefore, our initial strategy was to explore the prokaryotic MoeA-MogA complex assembly (separately expressed E- and G- domains compared to the fused eukaryotic Mo-insertases). This would be easier for studying the interaction between the domains and compare them with the other homologues Mo-insertase complexes in other species. For this purpose, I recombinantly expressed and purified MoeA and MogA separately, as well as co-expression of both. I applied various crosslinkers to stabilize the complex, isolate the complex, inspected the residues interacting from both proteins via XL-MS, investigated the complex size and stoichiometry using native-MS, negative staining to visualize the protein assembly, and modelled the complex using different computational modelling approaches. Furthermore, I complemented the prokaryotic MoeA-MogA with the plant Cnx1 separately-expressed domains to examine the differences in complex formation if present.

## 3.1.2. Purification and crosslinking of MoeA-MogA complex

To analyze the MoeA-MogA complex, I purified the proteins individually (as described in 2.2.2.1 & 2.2.2.2) and monitored the protein purity by SDS-PAGE, as shown in figure 10. Afterwards, the proteins were mixed in an equimolar ratio then loaded on a pre-calibrated gel filtration column. As a control, the individual proteins ran on the gel filtration column before running the mixture of the two proteins (Figure 13).

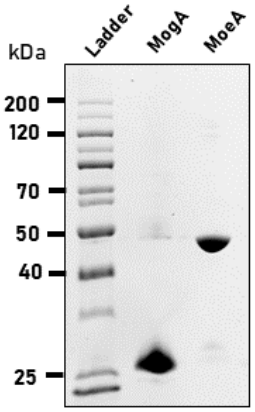



**Figure 13. SDS-PAGE of recombinant MogA and MoeA.** A 5 µg quantity of each recombinant protein was loaded onto the TGX stain-free SDS-PAGE. MogA is visualized with a size of ≈ 28 kDa and MoeA with a size of ≈ 45 kDa.

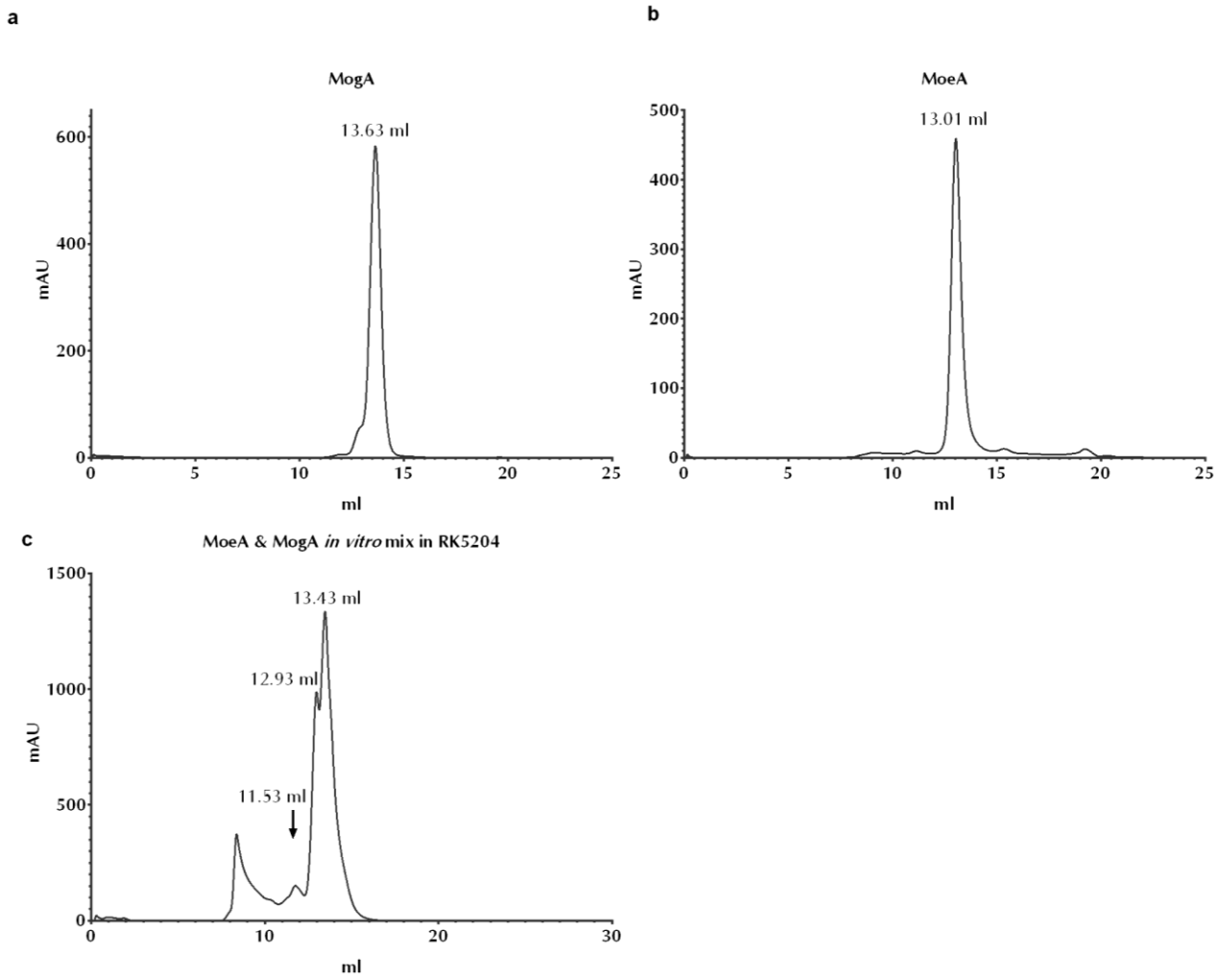



**Figure 14. Gel filtration of MoeA and MogA.** The proteins were loaded on a Superdex 200 increase analytical column. All proteins were eluted in 100 mM HEPES, 300 mM NaCl, 1 mM EDTA and 5% glycerol. (a) Chromatogram of purified MogA that elutes at 13.63 ml. (b) Chromatogram of purified MoeA which elutes at 13.01 ml. (c) *In vitro* mix of equimolar MogA and MoeA chromatogram showing complex elution at 11.53 ml, MoeA at 12.93 ml and MogA at 13.42 ml. All main peak fractions were tested with SDS PAGE.

Based on theoretical size calculation, which relies on the calibration of the gel filtration column, the MoeA-MogA complex correlates to 165.4 kDa complex size (approximate size of a trimeric MogA bound to a dimeric MoeA (Nichols & Rajagopalan, 2002). The main peaks were all analyzed with SDS-PAGE to ensure the correct identity of the proteins. In the gel filtration profile of MoeA, the prominent peak can be seen at 13.01 ml while 13.63 ml for the profile of MogA. For the mixed proteins, 3 peaks appear 11.53, 12.93 and 13.42 ml which correlates to MoeA-MogA complex, MoeA and MogA respectively (Figure 14). To exclude any influence on other factors on MoeA-MogA interaction, I also co-expressed both

proteins in *E. coli.* Here I used a strep-tagged MoeA as bait or strep-tagged MogA as bait (MogA or MoeA routine purification was possible by this approach). Thus I conclude that the *in vivo* co-expression of MoeA & MogA showed the identical band on the SDS-PAGE compared to the complex obtained from the *in vitro* mixture, regardless of the location of the StrepTagII® (C-terminus strep-tagged MoeA or C-terminus strep-tagged MogA) (Supplemental figure 68). As molybdate was shown to be essential to trigger MPT-AMP transfer from G- to E- domain (LIamas *et al.,* 2006), I added molybdate as well as the physio-chemical highly similar tungstate to my fully defined *in vitro* approaches. The complex peak was monitored under the influence of other conditions, including different pH and salt content. However, no observed difference in the complex compared to individual proteins in the chromatogram (Supplemental figure 74).

Interestingly, when comparing the RK5204 *E. coli* strains with protein purified from *E. coli* strain that lack the first step Moco biosynthesis proteins (knock-out *E. coli* strains KJW2-A (MoaA) and KJW3-A (MoaC) (K. Warnhoff and G. Ruvkun, 2019) (strains provided by T. Hercher)), complex formation was found to be depleted only in the chromatograms of strains that lack the first step Moco biosynthesis (see supplemental figure 74). Global landscape of protein complexes in *E. coli* using sequential peptide affinity (SPA) tags and MS showed an interaction of the first step of Moco biosynthesis (including MoaA and MoaC) with MoeA and MogA in a large protein complex with other proteins (Babu *et al.*, 2018). This hints that the complex in bacteria requires the first two proteins of Moco biosynthesis for the MoeA-MogA complex to be formed. My results further confirm this interpretation.

The results documented so far were obtained without employing any cross-linking reagent to stabilize the MoeA-MogA complex. In order to identify the interface(s) of the MoeA-MogA complex, in the following, I applied various cross-linkers both to the fully defined *in vitro* and *in vivo* approach. Initial *in vivo* cross-linking of MoeA and MogA was done using different cross-linkers, including formaldehyde, disuccinimidyl suberate (DSS), disuccinimidyl tartrate (DST), 1,5-difluoro-2,4-dinitrobenzene (DNDFB). Crosslinking efficiency was monitored using SDS-PAGE (Figure 15). When a mixture of DSS and DST (or 4% formaldehyde) were applied to the higher mixture of MoeA and MogA, a protein species of higher molecular weight (≈ 200 kDa) became detectable. A low molar excess of the

cross-linker (5-fold) did not show a significant cross-linked product. Nevertheless, the stable dimeric MoeA and trimeric MogA could be observed in the SDS PAGE upon crosslinking.

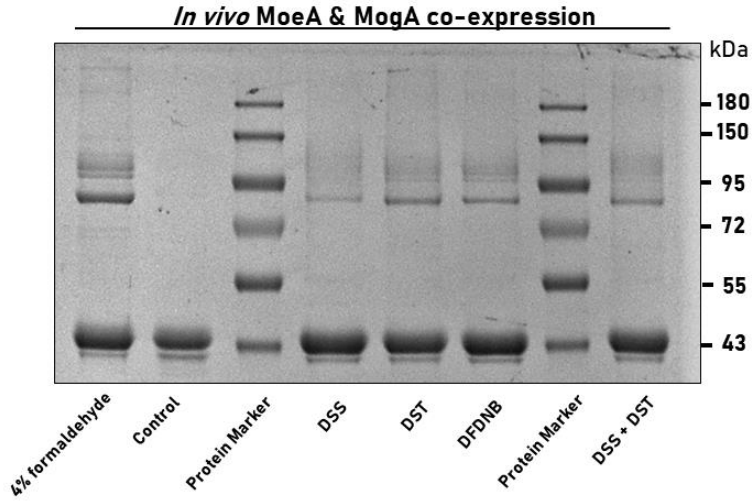



**Figure 15. SDS-PAGE of the initial attempt for the *in vivo* Crosslinking of MoeA and MogA with different cross-linkers.** *In vivo* cross-linking was done with a 5-fold molar excess of the cross-linker in the case of DSS, DST, and DFDNB. For DSS and DST, a total of 10-fold excess cross-linker was added and 4% formaldehyde. Protein complexes visualized as high molecular weight bands (more than 150 kDa) can be observed in the case of DSS + DST and the 4% formaldehyde. However, upon the application of DSS, DST, and DFDNB, no MoeA-MogA protein complexes were detected via SDS PAGE analysis.

A gradual increase of the molar excess of the cross-linker was tested to optimise the crosslinking conditions using DSS. As shown in figure 16, the increase of molar excess (up to 100-fold molar excess) resulted in the increase of the amount of the MoeA-MogA higher molecular weight complexes (≈ 200 kDa – 450 kDa). These can consist of a trimeric MogA in a complex with a dimeric MoeA. Correspondingly, other stoichiometries could be explained (e.g., two trimers and one dimer, 230 kDa), which were all theoretically calculated. However, to experimentally validate these results, other approaches, including MS, needs to be incorporated.

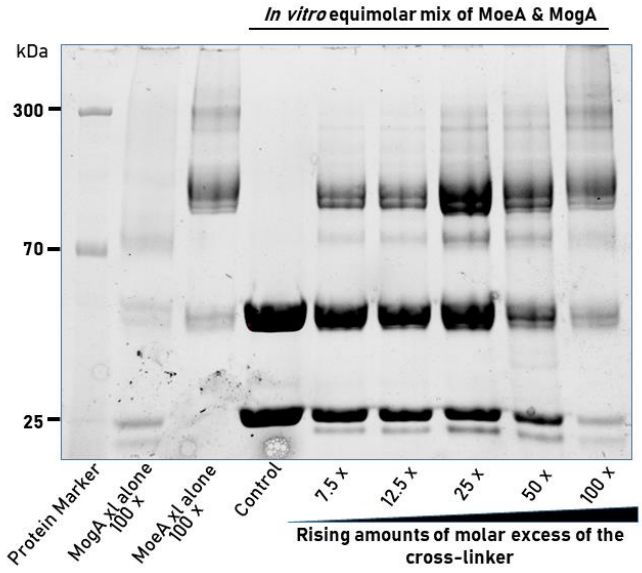



**Figure 16. SDS-PAGE of the rising amount of molar excess of the cross-linker for *in vitro* MoeA and MogA cross-linking.** DSS was used for all crosslinking reactions. A 12% SDS-PAGE visualized with TGX stain-free. MogA and MoeA were cross-linked individually with a 100-fold excess cross-linker in the second and third lane. As a control, MoeA and MogA equimolar mix was loaded on the SDS PAGE in the fourth lane (45 kDa and 23 kDa, respectively). Further, the increase in molar excess from 7.5 x molar excess to 100 x molar excess to investigate the impact of the cross-linker and determine the molar excess. At 7.5 x molar excess, only very faint high molecular weight bands can be observed. However, the more molar excess used, the more intense the high molecular bands are. Additionally, fading of the monomeric forms of MoeA and MogA which are recruited in the high molecular weight bands.

## 3.1.3. Native MS MoeA-MogA

Native-MS is a very sensitive technique for sizing complexes relying on electrospray ionization, which allows the exact determination of the molecular weight of proteins. Subsequently, complex stoichiometry could be deduced based on calculations (Natalello *et al.,* 2016). Importantly, native MS extract structural information of assemblies in their natural form, thus the term ''native'' (Leney & Heck, 2016). To understand more about the MoeA-MogA complex, I initially analyzed the separately expressed MoeA and MogA proteins as controls, allowing the visualization and size determination of each of their basic tertiary structure individually. The proteins were prepared in an appropriate total protein amount of ≈ 20 µM to avoid high concentrations (showing noisy signals) and low concentrations (showing no signal). For MogA, the distinct trimeric signal was observed with a molecular weight of 69 kDa with a charge state of +14 to + 17 (Figure 17a).

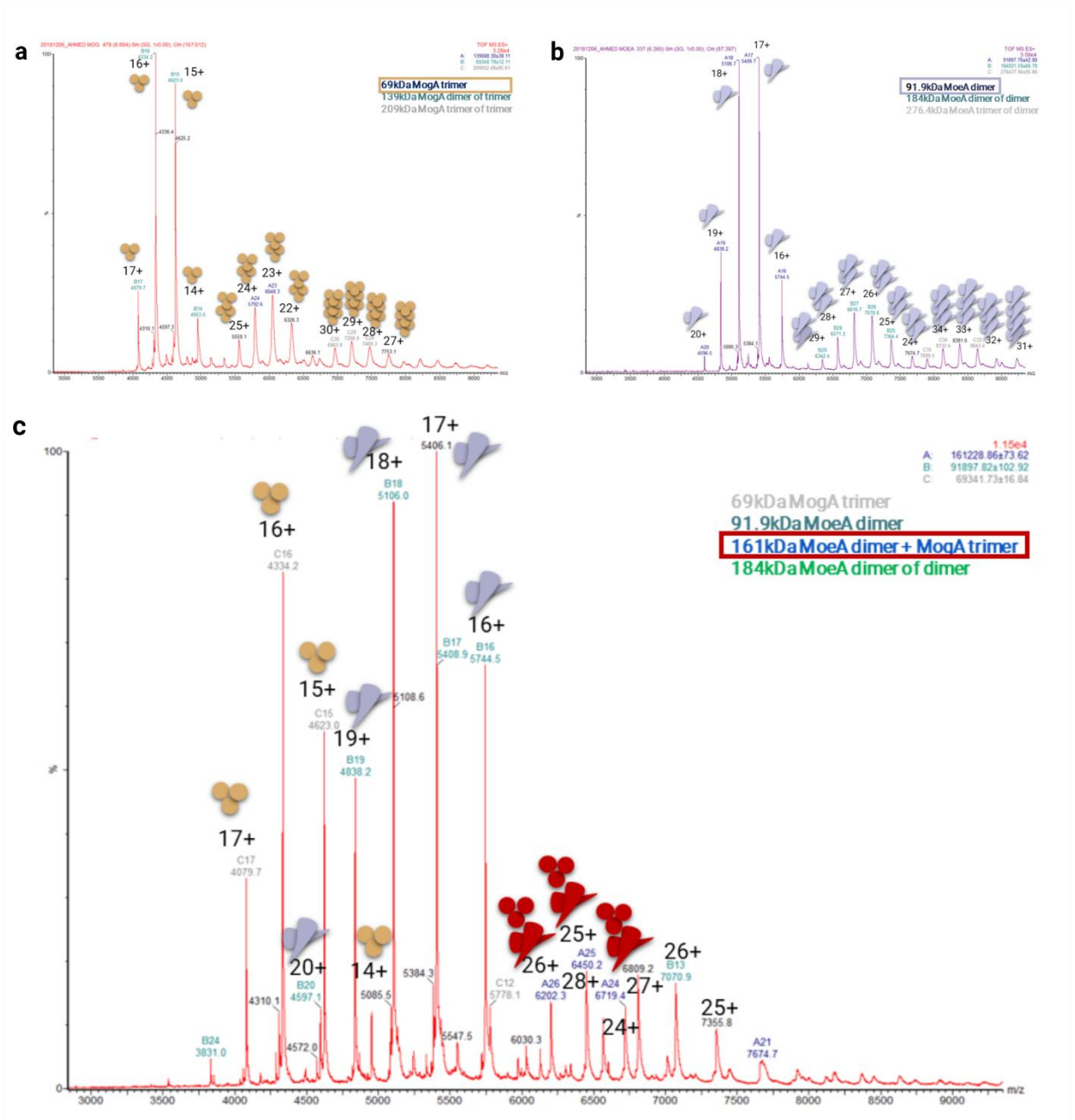



**Figure 17. Native mass spectrum of MogA, MoeA and MoeA-MogA *in vitro* Mix.** (a) Native spectrum of MogA showing trimeric species (charge state +14 to + 17) and native mass spectrum artifact of multimeric formation (charge state +25 to +30). The trimeric MogA is presented schematically above the designated peaks. (b) Native spectrum of MoeA showing a dimeric species (charge state +16 to +20) and native mass spectrum artifact of multimeric formation (charge state +25 to +34). The dimeric MoeA is presented schematically above the designated peaks. (c) Native spectrum of trimeric MogA (charge state +14 to +17), dimeric MoeA (charge state +16 to +20) and MogA-MoeA complex (charge state +24 to +28). The trimeric MogA, dimeric MoeA and the complex are presented schematically above the designated peaks.

Likewise, For MoeA, the mass spectrum showed clear signals of the MoeA dimer observed of a molecular weight of 91.9 kDa with a charge state of +16 to +20, which was identical

to the theoretically calculated molecular weight from the sequence (Figure 17b). However, in both of the cases of MoeA and MogA, tetramers and hexamers are observed after the prominent peak in the MS spectrum. The later minute peaks in the spectrum with charge state +25 to +34 or +25 to +30 are referred to as ''standard native MS artifacts''. This happens when the samples are sprayed from the non-denaturing solvent in the gas phase of the MS, causing some minor peaks to appear in the spectrum of the MS (Leney & Heck, 2016), therefore they can be ignored.

After incubating the equimolar mixture of both proteins for one hour, in order to allow equilibrium of complex formation, the samples were loaded on the native MS. The subsequently obtained MoeA-MogA spectrum showed three distinct species of a trimeric MogA, a dimeric MoeA and a heteromer of MoeA-MogA complex of size 69, 91.9 and 161 kDa, respectively (Figure 17c). At the same time, the charge state of the trimer and dimer (+14 to + 17 and +16 to +20, respectively) correlated to the charge states of the individual samples that ran as a control (Figure 17a and 17b). An additional, prominent charge state of +24 to +28 correlated to the complex of both MoeA and MogA. Due to the technical limitation of the equipment, a further resolution was not achievable. Even higher molecular weight species could also not be confirmed due to the instrumentation limits. Advanced instruments such as UHMR native MS can go for pseudo-MS 3 and measures up to 9.3 MDa (Keener *et al.,* 2021). Nonetheless, the molecular weight determination of proteins in all the native mass spectrometers is very sensitive with a margin error of approximately 10 Da (+ or -) (Vimer *et al.,* 2020). The MoeA dimer of 91.9 kDa and the MogA trimer of 69 kDa would add up to 160.9 kDa, which is the actual mass determined for the third visible peak of the complex of 161 kDa. Furthermore, as shown in table 22, I theoretically calculated Moco and Moco metabolites masses, including MPT, MPT-AMP, and Moco, to make sure that the mass difference of 100 or 200 Da does not correlate to a metabolite bound.

**Table 22. List of masses.** The table contains a list of masses of all Moco metabolites and MoeA-MogA protein combinations, including theoretical and native MS results.

| Entity | Theoretically calculated mass (Da) | Native MS results (Da) | Mass difference (compared to Native MS result) (Da) |
|---|---|---|---|
| MogA (Trimer) | 69331.02 | 69341.63 | 10.61 |
| MoeA (Dimer) | 91910.52 | 91887.97 | 22.55 |
| MogA (Trimer) + MoeA (Dimer) | 161241.54 | 161237.88 | 3.66 |
| MPT | 394.33 | | |
| MPT-AMP | 782.36 | | |
| Moco | 598.18 | | |
| Mo | 95.95 | | |
| Moco-AMP | 989.18 | | |
| MogA (Trimer) + 3 MPT | 70514.01 | | |
| MogA (Trimer) + 3 MPT-AMP | 71678.09 | | |
| MoeA (Dimer) + 2 Mo | 92102.42 | | |
| MoeA (Dimer) + 2 Moco | 93106.89 | | |

Notably, for all the proteins in the case of native MS, I expressed both proteins in the *E. coli* RK5204 strain, which lacks all metabolites of Moco (Rivers *et al.,* 1993).

Remarkably, in the field of native MS, it is still a matter of debate whether proteins are fully (100%) or mostly in the native form (90 to 99%) (Leney & Heck, 2016). The initial idea of partially preserved protein structure came while observing proteins that were electrosprayed from an aqueous solution. These proteins resembled narrower charge state distributions compared to proteins electrosprayed from an organic solvent (Vimer *et al.,* 2020). Partly maintained protein structures and not fully maintained, which are also due to the possible loss of ligands in the gas phase as an outcome (Leney & Heck, 2016). The previously expected theory was that the Mo-insertase complex is formed due to the presence of Moco precursors (Magalon *et al.*, 2002). Even though the proteins were expressed in a Moco metabolites-free strain, the complex assembly was still observed.

Importantly, this shows that the presence of MPT is possibly a reason to recruit more of the MoeA-MogA complex. However, it is not essential since the complex can be formed in the absence of all the Moco metabolites. As these experiments were conducted *in-vitro*, perhaps the key to enhancing complex assembly is still unknown, e.g., DNA or RNA. On the other hand, as briefly described in section 3.1.3, MoaA and MoaC can possibly be the initiators of complex formation in bacteria. Nevertheless, as a significant finding, the dimeric MoeA and trimeric MogA assembles into an asymmetric heteromer.

## 3.1.4. Crystallization of MoeA and MogA

For the crystallization of MoeA and MogA, I expressed and purified the proteins. The proteins were crystalized in collaboration with J. krausze. The structure of MoeA was firstly published by S. Xiang and co-workers in 2001 (Xiang *et al.,* 2001), while, MogA was published by Schwarz, G. in 2001 (Schwarz *et al.* 2001). MoeA and MogA were previously solved in 2.65 Å and 2.6 Å resolution, respectively. The crystal structure solved in this work of MoeA was at 2.1 Å resolution, while 1.9 Å resolution for MogA. A c alpha-based superimposition of the structures previously solved with the solved structures in this work is shown in (Figure 19b and 19c). At first, the crystallization of MoeA and MogA was done after mixing the proteins in an equimolar ratio as a trial to obtain a co-crystal structure of the MoeA- MogA complex. Later, all the crystals that grew and solved were crystals of the individual proteins in separate. Representative MoeA and MogA crystals are shown in figure 18.

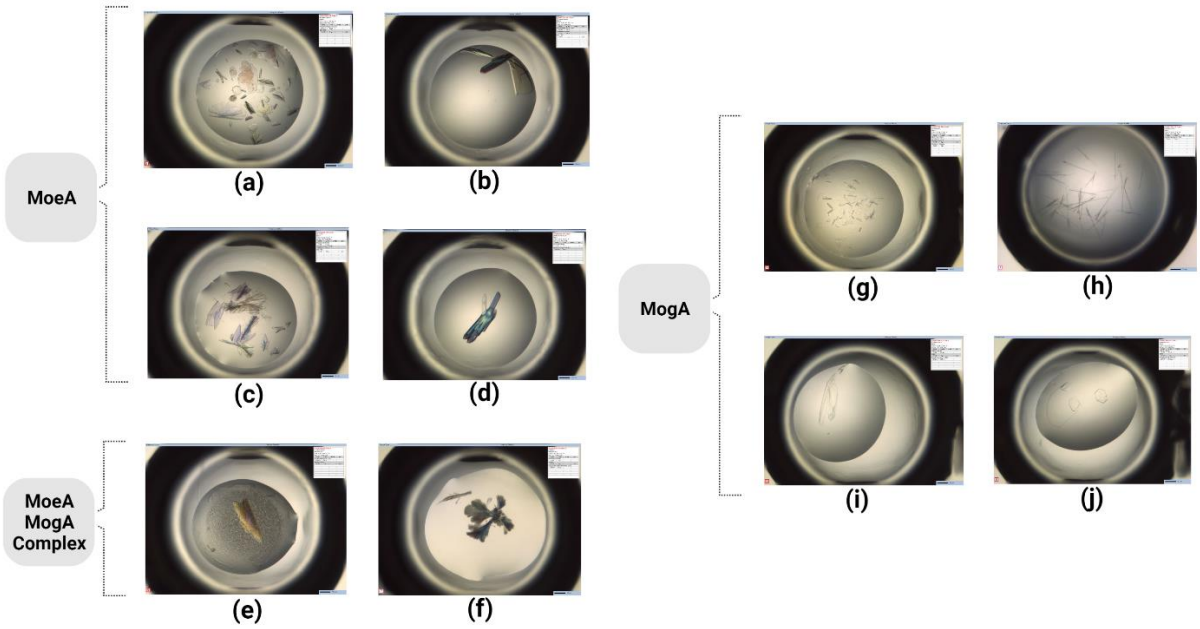



**Figure 18. Example of MoeA, MogA and MoeA-MogA complex crystals.** All crystals appeared after mixing 0.2 µL of the protein solution (concentration of ≈ 20 mg/ml) with 0.2 µL of the according reservoir. (a) Crystals obtained from 100 mM BIS-TRIS and 25% w/v PEG 3350. (b) Crystal growth resulted from 200 mM ammonium acetate, 200 mM BIS-TRIS and 25% w/v PEG 3350. (c) Crystals obtained from 200 mM $MgCl_2$, 100 mM BIS-TRIS and 25% w/v PEG 3350. (d) Crystal grew in 100 mM tri-sodium citrate and 20% w/v PEG 8000. (e) Crystal obtained from 100 mM sodium acetate and 25% w/v PEG 3350. (f) Crystals obtained from 200 mM ammonium sulfate, 100 mM HEPES and 35% w/v Poly(acrylic acid sodium salt) 2100. (g) Crystals acquired from 100 mM Morpheus buffer, 90 mM Morpheus halogens and 37.5% w/v MPD P1K P3350. (h) Crystals were in 8% PEG 4000. (i) Crystal obtained from 800 mM lithium chloride, 100 mM TRIS and 32% w/v PEG 4000. (j) Crystals grew in 500 mM lithium chloride, 100 mM TRIS and 28% w/v PEG 6000. (Crystallization was done with the help of J. Krausze)

Structural comparison of the MoeA structure reported here with the various E-domain structures available in the PDB revealed that the main difference was the hinge region of the flexible part of subdomain II and III. Using PDB Flex (Hrabe *et al.,* 2016), the solved MoeA structure showed the most extreme opening conformation compared to all E-domains crystal structures from the different species (Figure 19d). All E-domains from different crystal structures were compared using Conserved Domain (NCBI) (Shennan Lu *et al.,* 2020). For MogA, the structure showed similarity overall, with no notable difference (Figure 19b). All MoeA-MogA modelling and protein-protein docking was done using the crystal structures that were generated in this work.

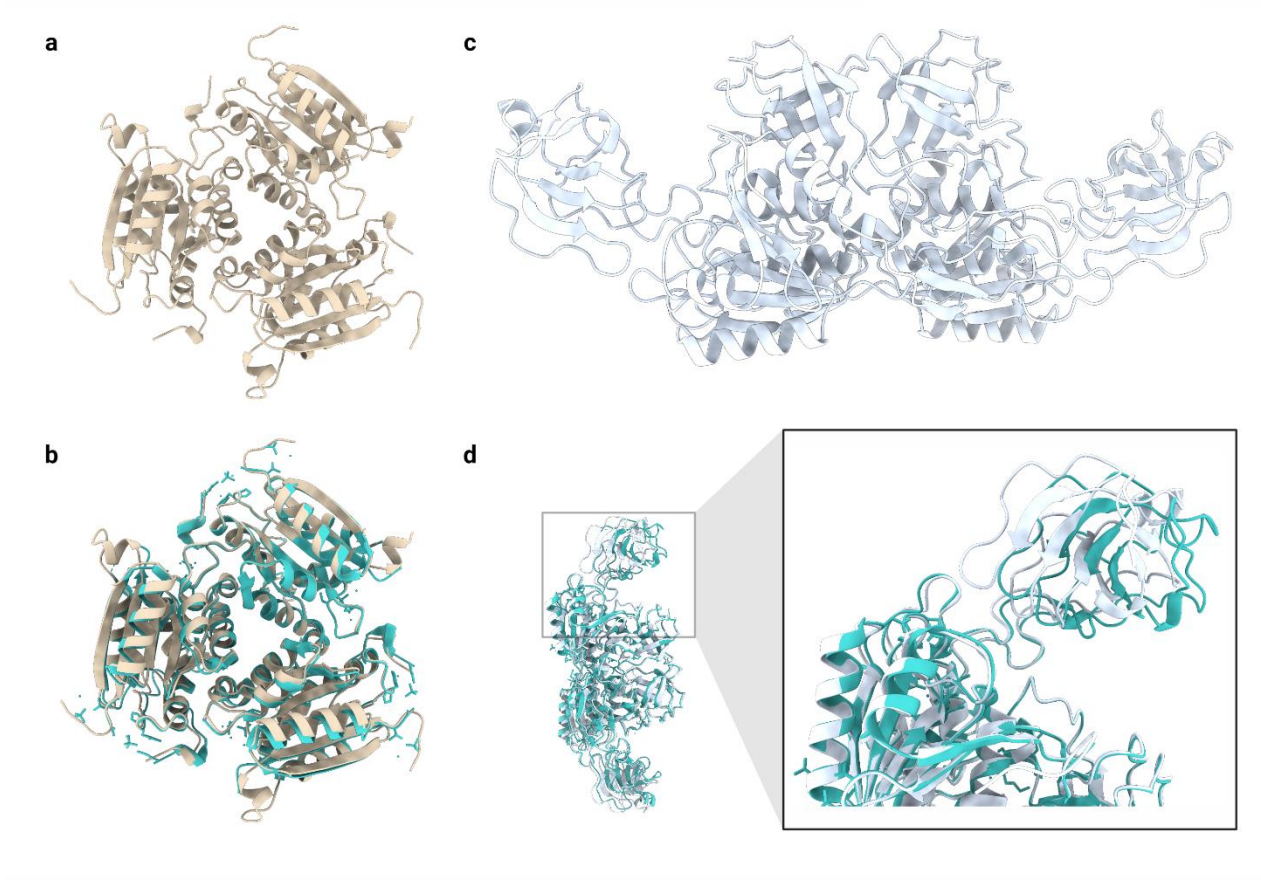



**Figure 19. Crystal structure of MoeA and MogA.** (a) Cartoon representation of MogA (this work). (b) Superimposition of MogA structure (this work, pale yellow) with the previously solved MogA structure (PDB: 1DI7) (cyan). (c) Cartoon representation of MoeA (this work). (d) Superimposition of MoeA (this work, light blue) with the previously solved structure of MoeA (PDB: 1G8R). The hinge region connecting subdomain II and III is shown enlarged with RMSD of 1.9 E (structures solved by J. Krausze).

## 3.1.5. XL-MS MoeA-MogA

The chemical cross-linking of MoeA-MogA was done *in vitro* (using an equimolar mixture of MoeA and MogA) and *in vivo* (using different *E. coli* strains which co-expresses MoeA and MogA). All XL-MS was done using the MS-cleavable disuccinimidyl sulfoxide (DSSO) (Developed by Huang group) or disuccinimidyl dibutyric urea (DSBU) (Developed by Sinz group) (Kao *et al.,* 2011; Ihling *et al.,* 2020). DSSO was used to optimize the cross-linking condition, whilst DSBU was used for the analysis of the cross-linked products. Prior to XL-MS experiments, the efficiency of XL was monitored by SDS PAGE revealing a decrease in the band of monomeric MoeA and MogA, and an enhanced band corresponding to the dimeric MoeA, trimeric MogA as well as multiple complex assemblies of a higher-order

structure (Figure 20). In order to confirm the presence of both proteins in the complexes, the higher molecular weight bands were excised from the SDS-PAGE gel as well as a control (a lower sized band), enzymatically digested them and injected the samples to the LC/MS/MS for analysis. The results from the MS analysis of DSSO cross-linked MoeA and MogA showed the presence of MoeA and MogA in all of the high molecular weight complex bands (Supplemental figure 69).

On the other hand, the control band excised of size approximate to 90 kDa showed only MoeA and the absence of MogA as expected. These results showed the formation of multiple complexes of MoeA-MogA, including the 161 kDa complex observed in the native MS (see 3.1.3). For DSBU, the procedure was identical. Nevertheless, different strains were employed for the cross-linking experiments (Figure 20), and the high molecular weight bands, as well as controls for LC/MS/MS, was excised.

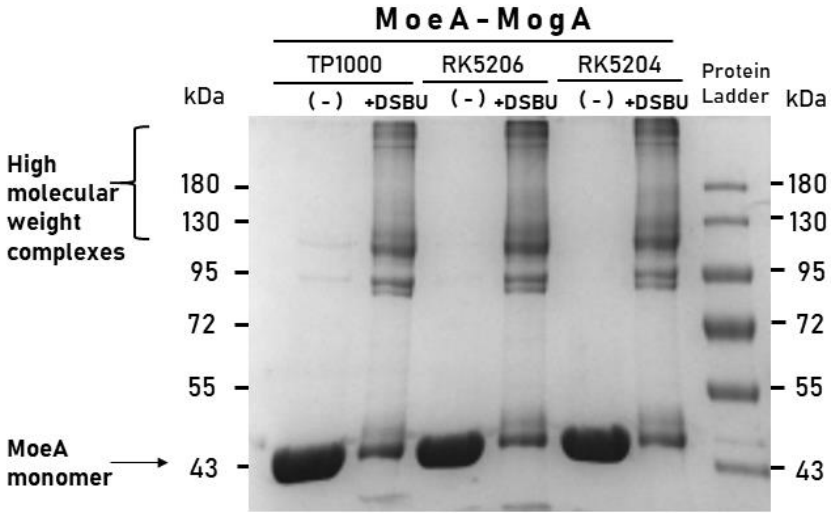



**Figure 20. SDS PAGE of MoeA-MogA *in vivo* crosslinking with DSBU in different *E. coli* backgrounds.** All cross-linking was done *in vivo* with DSBU in a 100-fold excess cross-linker to the proteins. The cross-linking was done using proteins expressed and purified from TP1000, RK5206 and RK5204 *E. coli* strains. However, no significant difference could be observed. The proteins were mixed in an equimolar ratio and incubated for an hour prior to crosslinking. The crosslinking showed high molecular weight bands (more than 130 kDa) with decreased intensity of the monomeric proteins. The high molecular weight bands were then excised, digested and subjected to LC-MS/MS.

Upon analysis with MeroX, chemical cross-linking of MoeA-MogA with DSBU showed plenty of intramolecular cross-links (within MoeA alone or MogA alone), which was insignificant for our purpose due to our knowledge of the crystal structures (see 3.1.4). Conversely, only two intermolecular cross-linked peptides (between MoeA and MogA) were highly reproducible across biological and technical replicates of all high molecular weight

complexes, including the ≈160 and ≈240 kDa complexes. The lysine residue number 140 (from the peptide SI**K**ETLEGVK of MogA; identifier number P0AF03) was cross-linked with lysine residue number 341 (from the peptide **K**TPGR of MoeA; identifier number P12281) as well as the lysine residue number 147 (from the peptide SIKETLEGV**K** of MogA; identifier number P0AF03) was cross-linked with lysine residue number 341 (from the peptide **K**TPGR of MoeA; identifier number P12281).

For chemical cross-linking, the biggest challenge is the analysis of the peptides in order to obtain accurate results with minimal false positives (Sinz, 2006). Due to the vast expansion of search space in the MS, unique peptides that ''stand out'' are advantageous in the analysis. MS-cleavable cross-linkers, which were initially introduced in 2001, such as DSBU, DSSO, and CDI, are superior for allowing easier and precise identification of cross-linked peptides as they show a unique fragmentation pattern in the gas phase of the MS (Ihling *et al.,* 2020). Although DSSO and DSBU are similar in the length of the linker and the chemical structure, DSBU is superior in comparison with DSSO because it does not require more than MS/MS to identify the two characteristics 26-u doublets in the spectra. On the other hand, DSSO requires identification under a third MS ($MS^3$), which is more complicated than the case of DSBU (Iacobucci *et al.,* 2018). It has to be noted that at first, the bands were digested with trypsin only. As a result, the peptides were too long to be analyzed since trypsin only cuts from the C-terminal of lysine and arginine (Dau *et al.,* 2019). Hence, another protease needed to be applied. For this purpose, *in silico* analysis of the digestion of MoeA and MogA with Trypsin and GluC (using ExPASy PeptideMass) was done. The results showed shorter peptide length when digesting with GluC and long peptides with trypsin (See supplemental). GluC, a serine protease that cuts glutamic and aspartic acid, was added overnight to ensure the short length of the digested peptides to allow the ease of analysis (Dau *et al.,* 2019). For the analysis, with MeroX (Sinz lab) software, the longer the peptide, the harder it is for the software to discriminate between actual cross-linked peptides and modified peptides with others.

Another hurdle for the MoeA-MogA analysis was the keratin, an abundant essential protein for hair, skin and nails, and contamination during the first XL-MS attempt, resulting in keratin coverage nearly as high as the coverage of MoeA and MogA. Consequently, the analysis was terminated (Supplemental figure 65). All band excision was performed under the fume hood with lab safety sleeves and gloves on to avoid this problem.

Although the amine-to-amine cross-linker used interacted with amine-reactive residues (Sinz *et al.*, 2015), the presence of the same inter-crosslink in MoeA-MogA samples, both *in vivo* and *in vitro*, can be only explained that the only amine-reactive residue available, in the interaction surface of both MoeA and MogA, is the merely one identified by the software. These results can be further elucidated by modelling the MoeA-MogA complex, with the knowledge that the interface between both proteins should not contain an amine-reactive residue in close distance to the identified ones.

## 3.1.6. Modelling the MoeA-MogA complex

For modelling the MoeA-MogA complex, two modelling approaches were done, including global docking (modelling MoeA - MogA docking without any crosslinking guidance) and guided-based modelling (using crosslinking restraints). For MoeA-MogA docking, ClusPro was used as a server for docking the two directly interacting proteins (Kozakov *et al.,* 2017) (see 2.2.7.2.). Following the steps mentioned in 2.2.7, the solved MoeA and MogA structures were docked, resulting in 30 models. The resulted models were a balanced set of electrostatic-favoured, hydrophobic-favoured and van der Waals electrostatics (Figure 21). All the results generated resembled an inverse ''L'' shaped complex of MoeA & MogA.

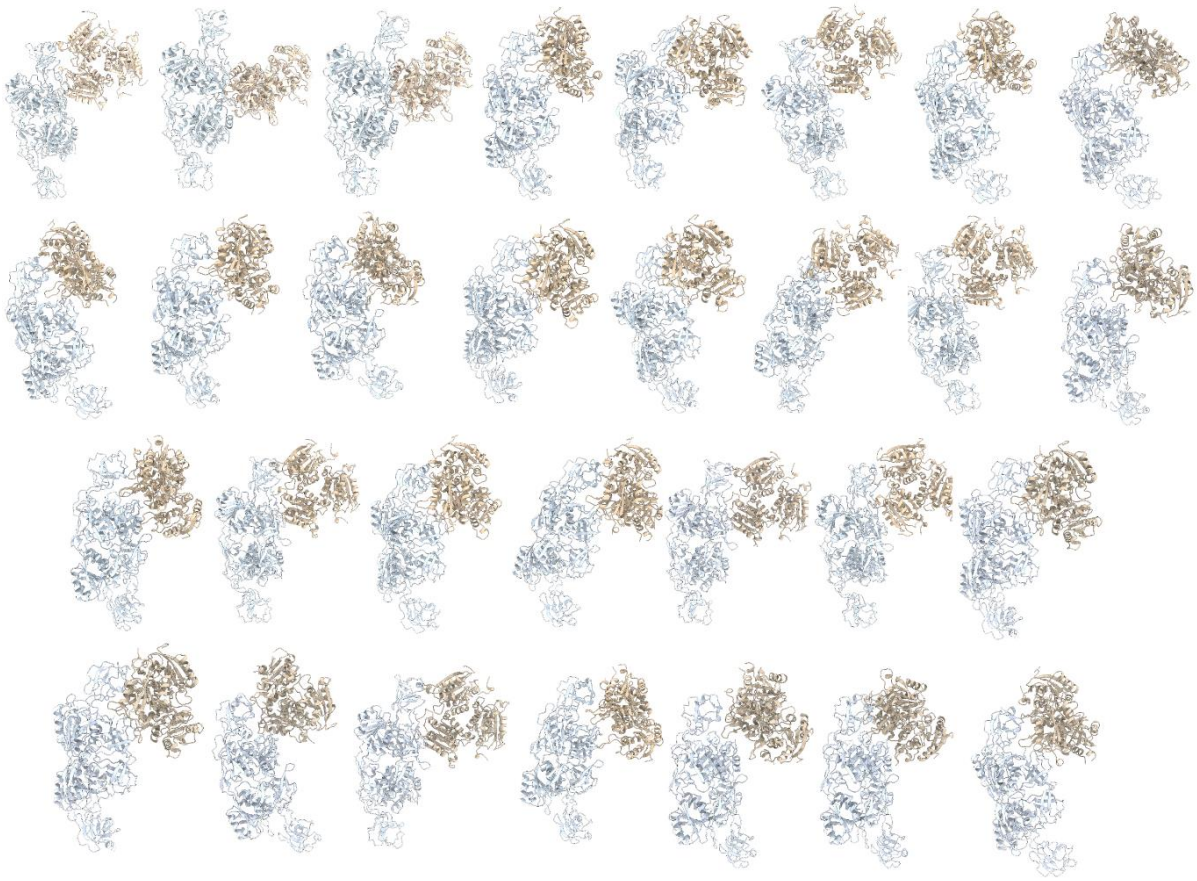


**Figure 21. Global docking of MoeA and MogA.** Thirty models were generated using ClusPro. The structures of MoeA and MogA in this work was used as a receptor (MoeA) and ligand (MogA) for the docking approach. MoeA is shown in light blue and MogA in pale yellow. The resulting models are based on a balance of electrostatic-favoured, hydrophobic-favoured and van der Waals electrostatics

Next to the non-guided docking, the two identified cross-linked MoeA-MogA peptides were subsequently used as restrains in a guided docking using HADDOCK (Koukos *et al.*, 2020). HADDOCK workflow presented in this work is shown in figure 12. The solved MoeA and MogA structures were used as input files for modelling after editing them using Coot and PHENIX (Emsley *et al.,* 2009; Liebschner *et al.,* 2019) (see 2.2.7.3). The final model is presented in figure 22. TBL files (files containing the information required for modelling) were prepared with only two restraints of residue 140 and 147 of MogA cross-linked with residue 341 from MoeA highlighted in figure 22b.

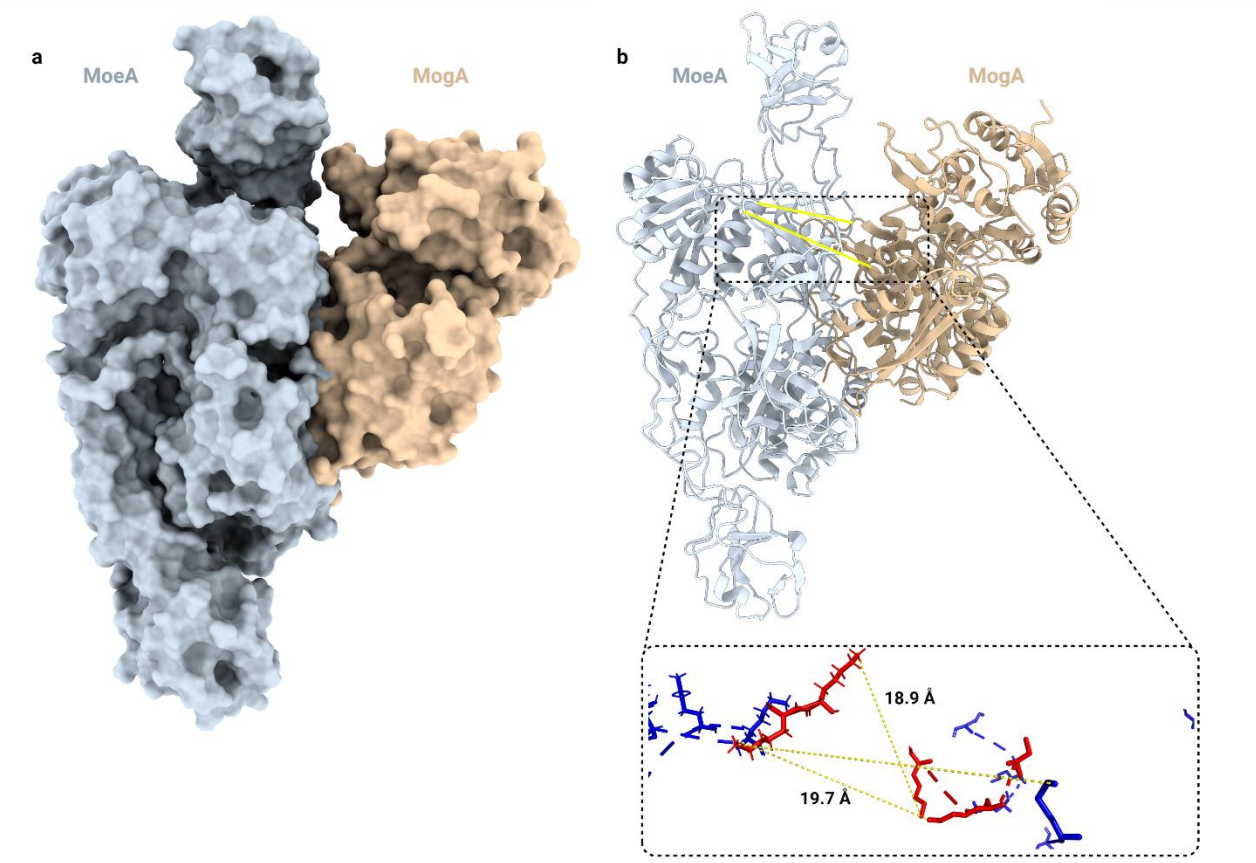



**Figure 22. MoeA-MogA complex model based on restraints.** (a) Surface representation of the restraint model generated using HADDOCK, MoeA is represented in light blue and MogA in pale yellow. The structures in the model used were the crystal structures solved in this work (figure 19). (b) Cartoon representation of the HADDOCK model highlighting the restrains used for modelling in yellow. The residues cross-linked are within the 18.9 and 19.7 Å range (red), all other lysine residues are represented in blue, and they all are within more than 28 Å distance as displayed in a magnified view of the restraints zone that is framed in a dashed line.

Importantly, with the exception of the two lysine residues that are highlighted in figure 22b, the lysine surface distribution on the model did not show any lysine residues in close proximity (the allowed distance of the cross-linker to interact with two residues). Moreover, the distance of the closest lysine residues in-between MoeA-MogA (excluding the identified cross-links) in the current model is 37 Å, which is beyond the span of the linker. The global docking approach was finally compared to the restraints-guided approach, and the results showed high similarity to where the MogA would bind to MoeA.

### 3.1.7. Negative staining MoeA-MogA complex

Negative staining of MoeA-MogA complex, which was purified from the chromatogram shown in figure 14, showed relative heterogeneity. Although the samples were slightly crowded, L-shaped particles are observed in the micrographs (Figure 23a, 23b and 23c). 2D classification of the particles showed classes that resembled an elongated or inverse ''L''-shape. These findings, in turn, are similar to the results coming from the global docking (Figure 21) and the crosslinking-based model (Figure 22). The organization of the E- and G-domain are derived from the results obtained from the 2D classes is shown in figure 23f. Consistently findings from the 2D classification are in complete agreement with results from gel filtration and native MS.

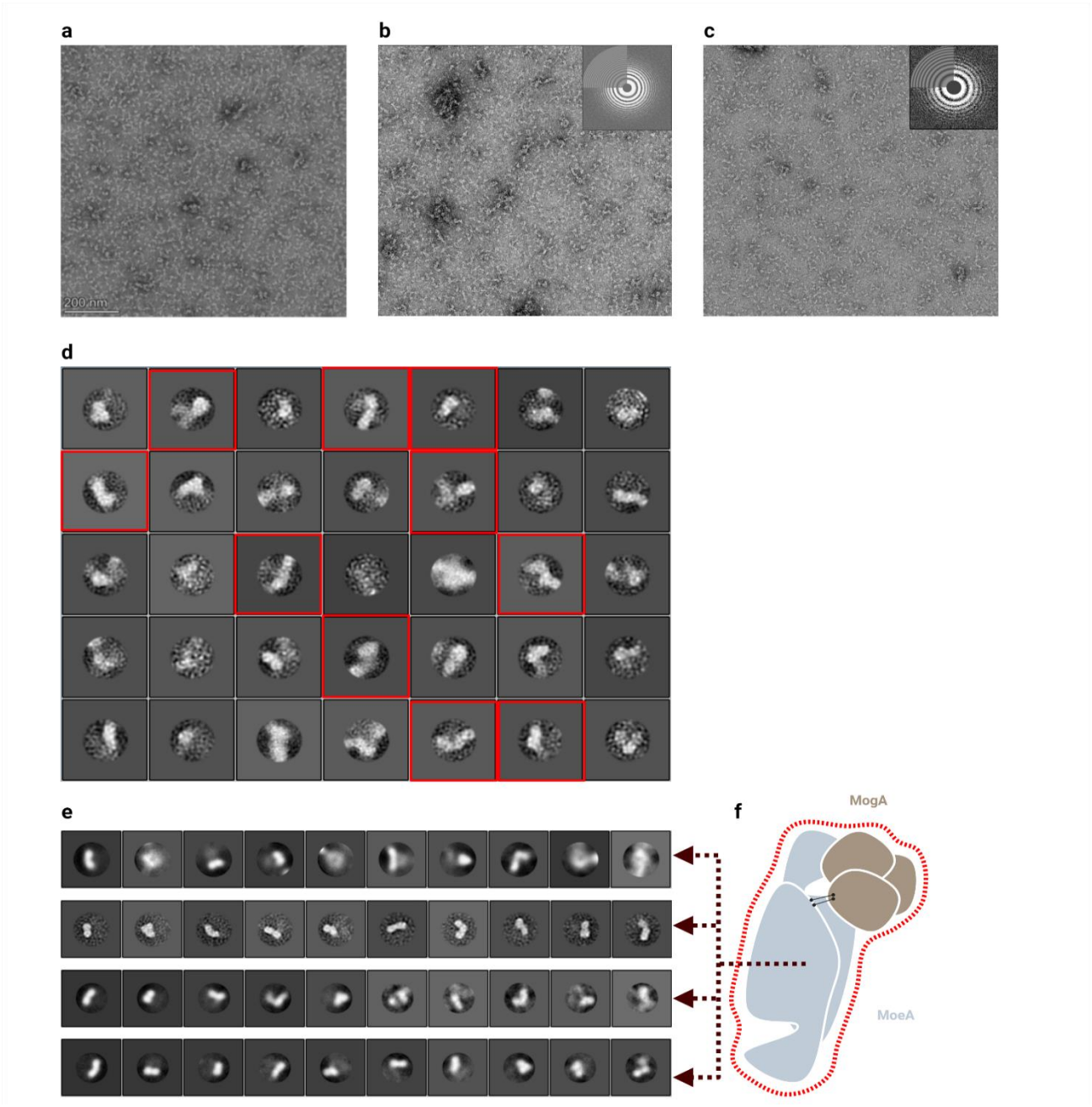



**Figure 23. Negative staining of MoeA-MogA complex.** All samples were cross-linked in 100 mM HEPES, 300 mM NaCl and 1mM EDTA and have a particle size of 90 to 150 E (a) Negative staining of MoeA-MogA large complex (e.g. two MoeA dimers and 2 MogA trimers) of the expected size of 250 to 300 kDa. The analyzed samples 8100 µg/µl) showed a relatively high degree of heterogeneity. (b) First negative staining of MoeA-MogA complex (trimer-dimer) of the expected size of 161 kDa. The samples were in 200 µl/ml concentration and relatively homogenous. 102 micrographs were taken, and approximately 70000 particles were used for auto picking. (c) Second negative staining of MoeA-MogA complex (trimer-dimer) of the expected size of 161 kDa. The samples were in 100 µl/ml concentration. 47 micrographs were taken, and approximately 24000 particles were used for auto picking. (d) Autopicking of the first negative staining micrographs. 192 boxsize, 250 E mask used and 100 classes were selected. Selected classes used as a trial for further classification are highlighted in red. (e) Autopicking and classes of second negative staining done. 192 boxsize, 250 E mask used and 5140 particles selected. Single particle picking was possible, and it was visually possible to see the inverted ''L''-shaped particles as well as the elongated shapes. (f) Schematic representation of MoeA and MogA model, which shares a mutual resemblance to the single particles picked (Negative staining EM results obtained and processed with the help of T. Beck and P. Wendler).

### 3.1.8. Complementation of prokaryotic Mo-insertases complex with plant Mo-insertase domains

It is previously shown that gephyrin reconstitutes Moco biosynthesis in prokaryotes and plants (Schwarz *et al.,* 1999). Nevertheless, the protein complex assembly of Mo-insertases was not shown to be complemented formerly. In the following, I tested for the ability of Cnx1E and Cnx1G to complement MoeA and MogA within the Mo-insertase complex. Therefore, the following co-expression constructs were created: Cnx1E-MogA and MoeA-Cnx1G as well as Cnx1E-Cnx1G and MoeA-MogA. All constructs were expressed in *E. coli* strain RK5204. For subsequent crosslinking, a 100-fold molar excess of DSSO was applied. The cross-linked products were then analyzed via SDS-PAGE (Figure 24a) and native gel electrophoresis (Figure 24b). To confirm the identity of the cross-linked products, an anti-step tag antibody (Ringel *et al.,* 2013) was used for immunoblot-based detection. As documented in figure 24a, the formation of Mo-insertase protein complexes was traceable for all expression strains analyzed. Remarkably, the documented sizes of the different complexes identified are highly comparable to that of the MoeA-MogA co-expression strain serving as control here (Figure 24a). On the other hand, analysis via native gel electrophoresis revealed differences amongst the expression strains analyzed. Here the Cnx1E-MogA co-expressing strains were found to possess a relatively more Mo-insertase complex as estimated from the band intensity (size ≈ 235 kDa) compared to the control (MoeA-MogA co-expression strain).

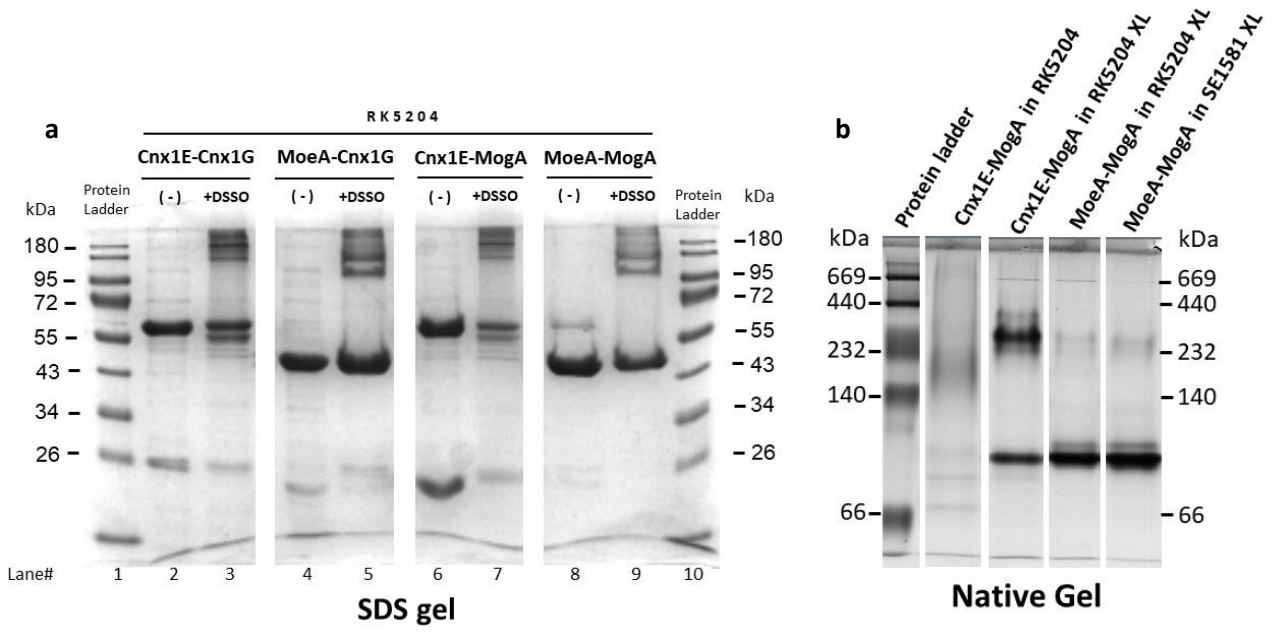



**Figure 24. SDS PAGE and native gel of complementing G- and E- domains from plants and bacteria.** All proteins were co-expressed in RK5204 and cross-linked *in vivo* with 100-fold excess DSSO cross-linker. (a) SDS gel shows that the respective strains were used for an *in vivo* crosslinking approach upon co-expression of E- and G- domains. Here, the 100-fold excess (+) was used. As a control, no cross-linker was applied (-). (b) Native gel using crude extracts of the analyzed strains in (a). Prior to analysis, the indicated strains were used for an *in vivo* crosslinking approach using the 100-fold excess DSSO. Additionally, MoeA-MogA were cross-linked upon expression in *E. coli* strain SE 1581.

The results indicate that the Mo-insertase complex formation occurs even when the E- and G domain of different species are involved. However, the results obtained require further confirmation, e.g. native MS-based, XL-MS, and so forth.

### 3.1.9 Summary

In summary, a MoeA-MogA complex of ≈ 161 kDa size was identified and characterized in this work by employing different techniques, including native MS, XL-MS, negative staining and computational modelling. For modelling, two approaches (XL-MS based docking and global docking) have been employed, which both yielded essentially the same results. Each of the approaches was found to be consistent with the negative staining EM approach. Therefore, four lines of evidence identified a first MoeA-MogA interaction model.

## 3.2. Cnx1 full-length complex

### 3.2.1. Overview

In Cnx1, the G- and E- domains are connected via a short linker of 23 amino acid residues (Schwarz *et al.,* 2000). However, since evolutionarily preserved proteins tend to form homologues complexes (Landry *et al.,* 2013), it can be assumed that the prokaryotic E- and G-domain form a comparable protein complex assembly in comparison with plants. If not identical, minor movements are expected of the domains when comparing different Mo-insertase complexes from different species. For Cnx1, I recombinantly expressed and purified the full-length protein, applied XL-MS, modelled the complex as well as negative staining and Cryo-EM to visualize the complex assembly of Cnx1. A summary of Cnx1 XL-MS and modelling is shown in figure 25.

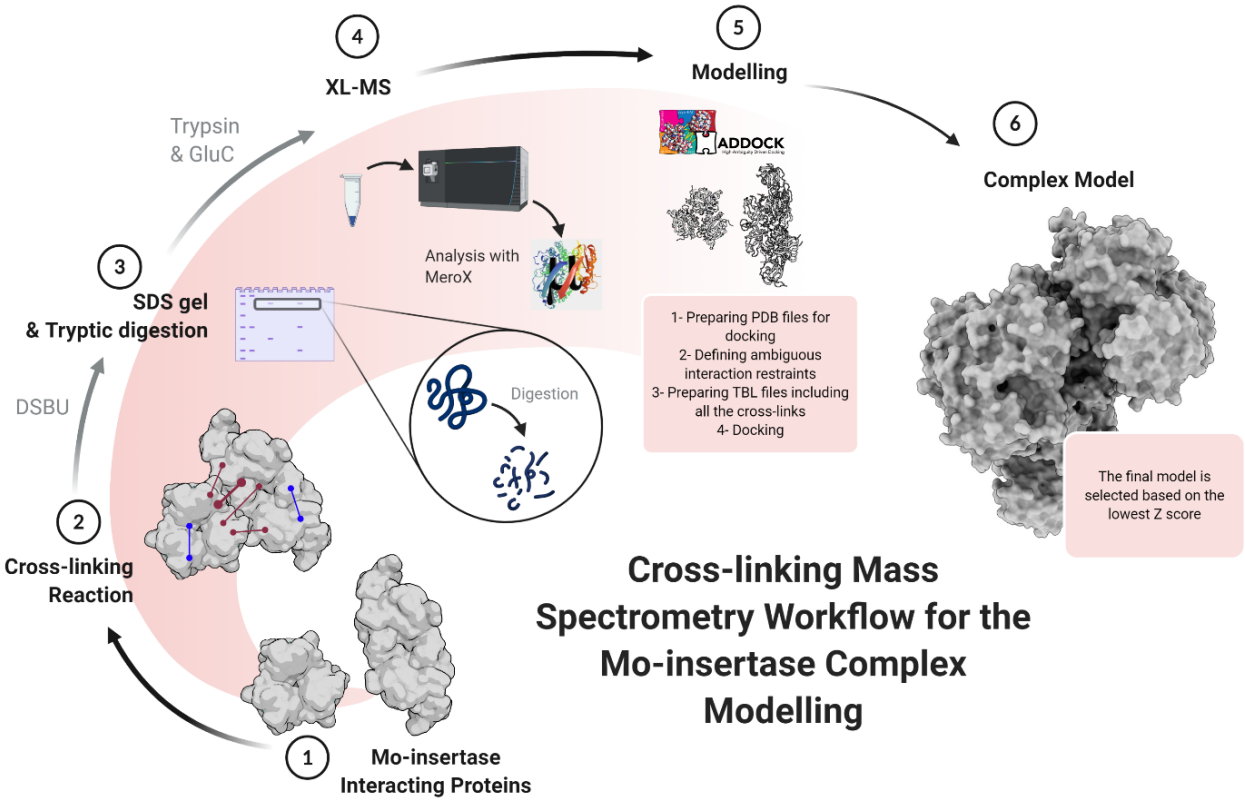



**Figure 25. Cross-linking Mass spectrometry workflow for Cnx1 complex modelling.**

## 3.2.2. Expression and purification of Cnx1 full-length

The recombinant expression and purification of full-length eukaryotic proteins are not often feasible and, in some cases, unachievable in *E. coli*. This is due to several factors, including unsuitable expression and purification conditions as well as proteases activity leading to low protein yield and/or the vast majority of the protein to be degraded (Rosano & Ceccarelli 2014). To establish a Cnx1 purification protocol that allows the production of (non-degraded) full-length protein, I initially expressed and purified the protein as C-terminal StrepII Tag fusion, following the parameters established for MoeA / MogA expression and purification (See 2.2.2.1 and 2.2.2.2). Nevertheless, doing so yielded protein preparations containing significant amounts of degraded Cnx1 full-length protein (Figure 26).

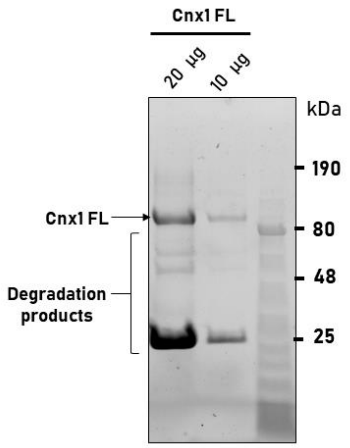



**Figure 26 SDS PAGE of the initial purification of Cnx1 full-length.** The proteins were loaded in 20 µg and 10 µg amount for comparison. Full-length Cnx1 can be observed at approximately 90 kDa and plenty of degradation products can be shown below 80 kDa.

In order to further enhance the homogeneity of the obtained protein preparation, I introduced an additional His6-tag on the C-terminus of the protein. However, the SDS-PAGE showed degradation product regardless of the double-tag purification (on both C- and N- terminus) of the protein. Therefore, subsequent downstream purification steps were included in the protocol (Figure 27). The inevitable presence of degradation products could be explained by the binding of the degradation products (fragments of G- and E- domain) with the full-length protein leading to the trapping of these products in a complex that can be co-purified by affinity chromatography.

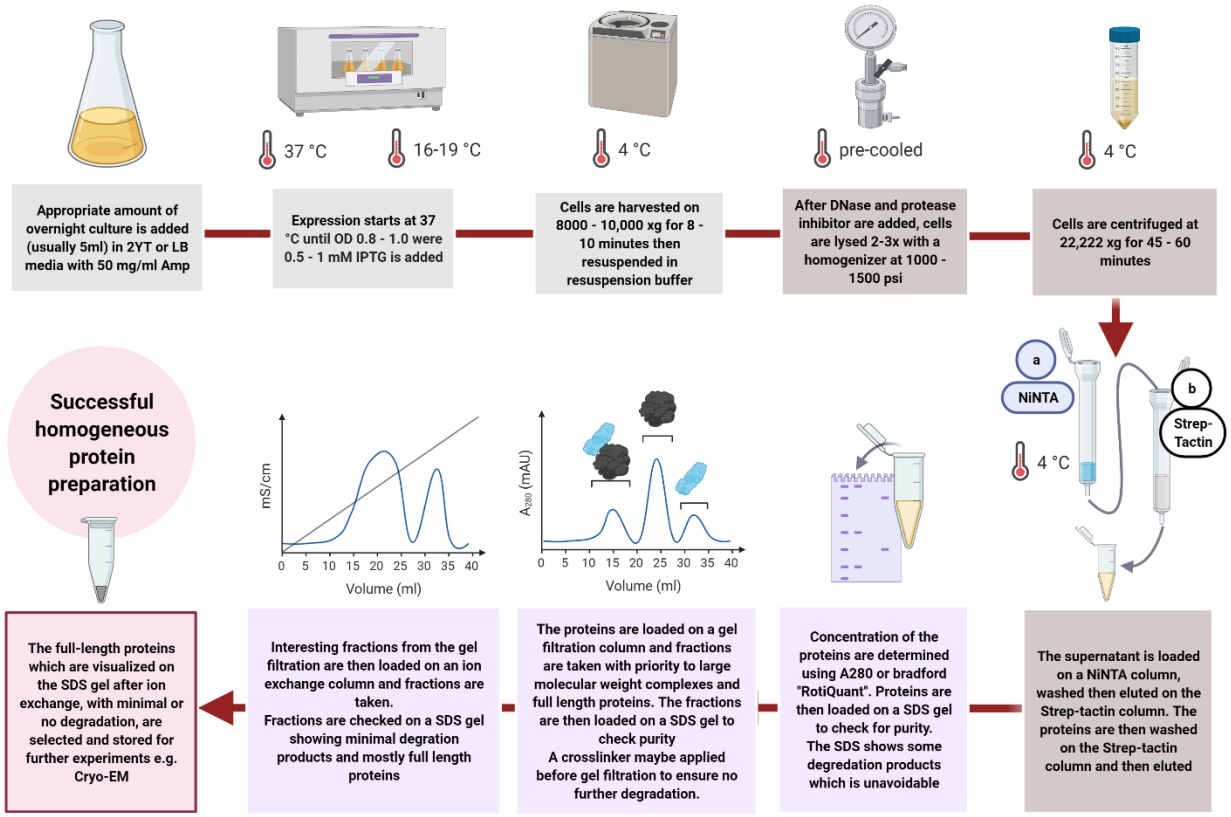



**Figure 27. Summary of the new expression and purification protocol of full-length proteins used in this work.**

Precisely, upon two-step affinity tag-based purification, the obtained protein preparations were subsequently purified by gel filtration (Figure 28 and 29). and anion exchange chromatography (Figure 30 and 31)

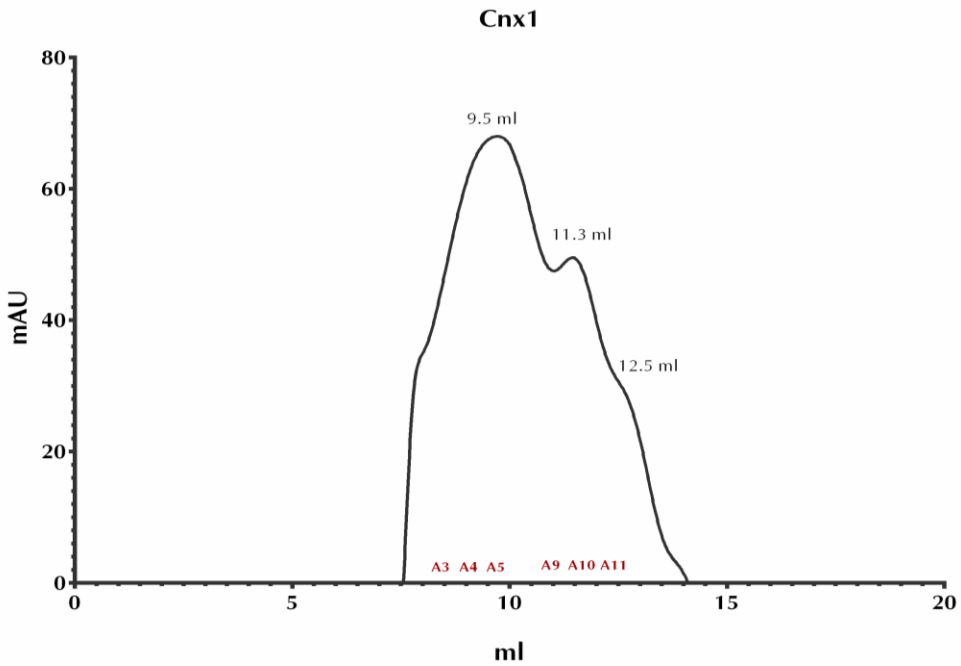



**Figure 28. Gel filtration chromatogram from the Cnx1 full-length purification.** The proteins were loaded on a Superdex 200 increase analytical column. All proteins were eluted in 100 mM HEPES, 300 mM NaCl, 1 mM EDTA and 5% glycerol. The fractions eluted for further SDS-PAGE analysis are labelled A3, A4, A5, A9, A10 and A11.

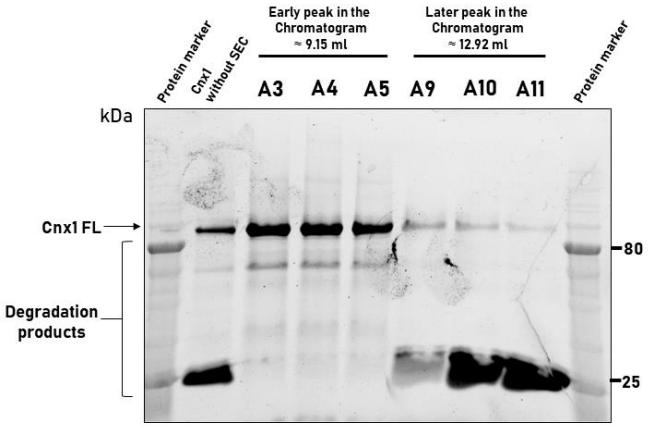



**Figure 29. SDS PAGE of the gel filtration fractions taken from the Cnx1 full-length purification.** A control sample of unpurified Cnx1 was loaded as a comparison. Fractions A3 to A11 were taken from the gel filtration run with the chromatogram shown in figure 28. The two main peaks of the Ion-exchange chromatogram were loaded on the gel, the early peak from the chromatogram was from A3 to A5, while the late peak was from A9 to A11. Fraction A3, A4 and A5 showed more full length compared to fraction A9, A10 and A11. Therefore, A3, A4 and A5 fractions were pooled and used for further ion-exchange chromatography step.

Afterwards, the interesting fractions A1 to A5 were pooled and used for the subsequent IEX. Prior to IEX, the proteins were cross-linked with 100-fold excess DSS cross-linker. The cross-linked protein samples were loaded on the IEX column with an 80 min salt gradient as described in section 2.2.2.2.5, and all fractions were collected. The IEX chromatogram

showed two distinct peaks appearing from a conductivity of 20 mS/cm (Figure 30). The fractions were collected and checked on SDS-PAGE (Figure 31). The late fractions of the first peak of the chromatogram showed degradation, yet, the early peak in the chromatogram showed full-length proteins and almost abolished degradation products.

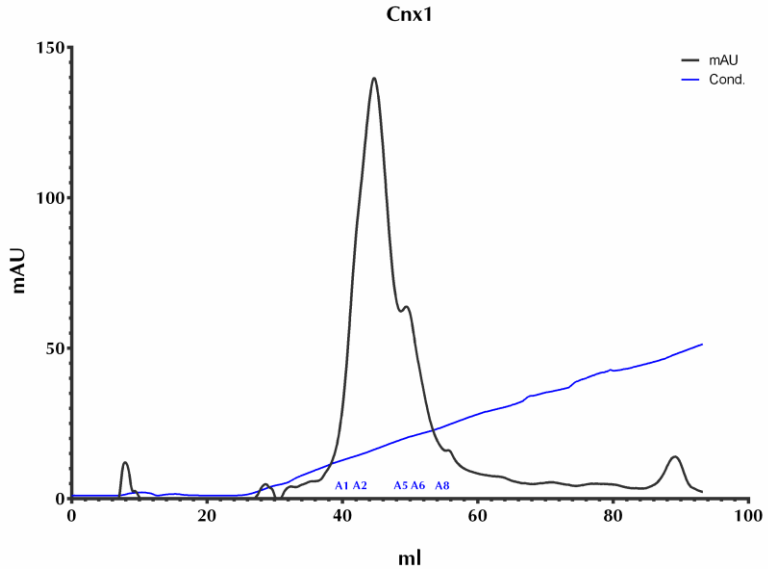



**Figure 30. Ion exchange chromatogram from the Cnx1 full-length purification.** The proteins were loaded on a Source Q15 analytical column. The 80 min gradient started with 20 mM Tris-HCl buffer and eluted with 20 mM Tris-HCl and 1M NaCl. The fractions eluted for further SDS-PAGE analysis are labelled A1, A2, A5, A6, and A8.

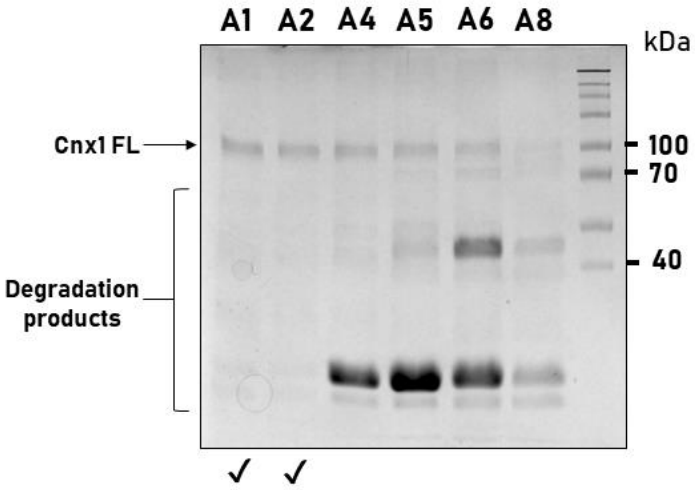



**Figure 31. SDS PAGE of the ion-exchange fractions taken from the Cnx1 full-length purification.** Fractions from A1 to A8 were obtained from the ion-exchange with the chromatogram shown in figure 30. Fraction A4, A5, A6 and A8 showed significant degradation products as well as less amount of full-length proteins. The Cnx1 full-length is of an approximate size of 90 kDa, while band A6 and A8 show a size of approximately 45 kDa, which is equivalent to the E-domain. In A4, A5, A6 and A8, bands of G-domain approximate size can be seen. Finally, A1 and A2 show no degradation product demonstrating the successful purification of the full-length protein (labelled by a checkmark).

Upon successful establishing a purification regime that allowed for the purification of recombinant full length Cnx1 without any significant degradation products, unless stated otherwise, thus obtained protein was used for crosslinking and negative staining EM experiments. In some cases, for significantly homogenous sample preparation for the EM, a cross-linker was added before or after gel filtration to ensure an intact complex and avoid dissociation over time.

## 3.2.3. XL-MS of full-length Cnx1

Chemical crosslinking of Cnx1 full length was performed with the DSBU cross-linker described in (2.2.4). Checking the protein complex on a SDS-PAGE after cross-linking is crucial to evaluate the correct protein folding, which can be distorted by over-cross-linking (Yu & Huang 2018), especially in the case of highly flexible proteins such as Cnx1. Therefore, SDS-PAGE was used, which documented the DSBU-dependent formation of two high molecular weight complex (≈ 200 kDa and ≥ 300 kDa). The corresponding protein bands were excised from the gel, enzymatically digested, and then used for the LC/MS/MS

analysis. Using MeroX to identify the cross-linked amino acid residues, all the cross-linked products were manually checked (Iacobucci *et al.,* 2018).

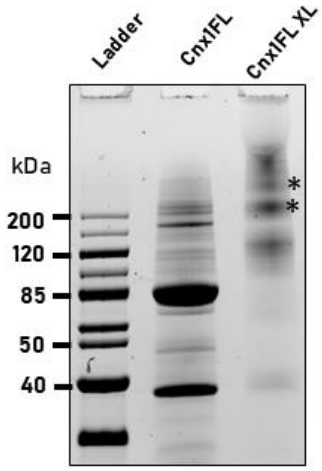



**Figure 32. SDS-PAGE of crosslinking Cnx1 with DSBU for LC-MS/MS.** A 12% SDS-PAGE visualized with TGX stain-free of crosslinking Cnx1 full-length. DSBU was used for the crosslinking reactions (Cnx1FL XL). High molecular weight complexes of ≥200 kDa is observed in the cross-linked sample compared to the control sample with full-length protein at ≈ 90 kDa (Cnx1FL). The high molecular weight complexes (marked with asterisks) were excised, enzymatically digested and subjected to LC-MS/MS.

As briefly discussed in 3.1.5, MS-cleavable cross-linkers are cleaved at their labile bonds in the MS/MS. This results in distinctive fragment ion patterns revealing a cross-link. Subsequently, MS 3 experiments are performed on the reporter ions of the cross-linker. CID (collision-induced dissociation) and ETD (electron-transfer dissociation) experiments could be done as an alternative by combining information from two complementary approaches of fragmentation (Iacobucci *et al.,* 2018). For DSBU, neither MS 3 nor CID and ETD experiments are necessary to be conducted, which is advantageous for this cross-linker. Unlike other MS-cleavable cross-linkers, the labile bonds of DSBU are cleaved at the same energy regime necessary to cleave the peptide backbone, allowing the detection of the unique fragments of the peptide and the linker at the MS/MS level (Ihling *et al.,* 2020) (Figure 32). Accordingly, this gives information about both the cross-linking site and the amino acid sequences of the peptides connected in one collisional activation step.

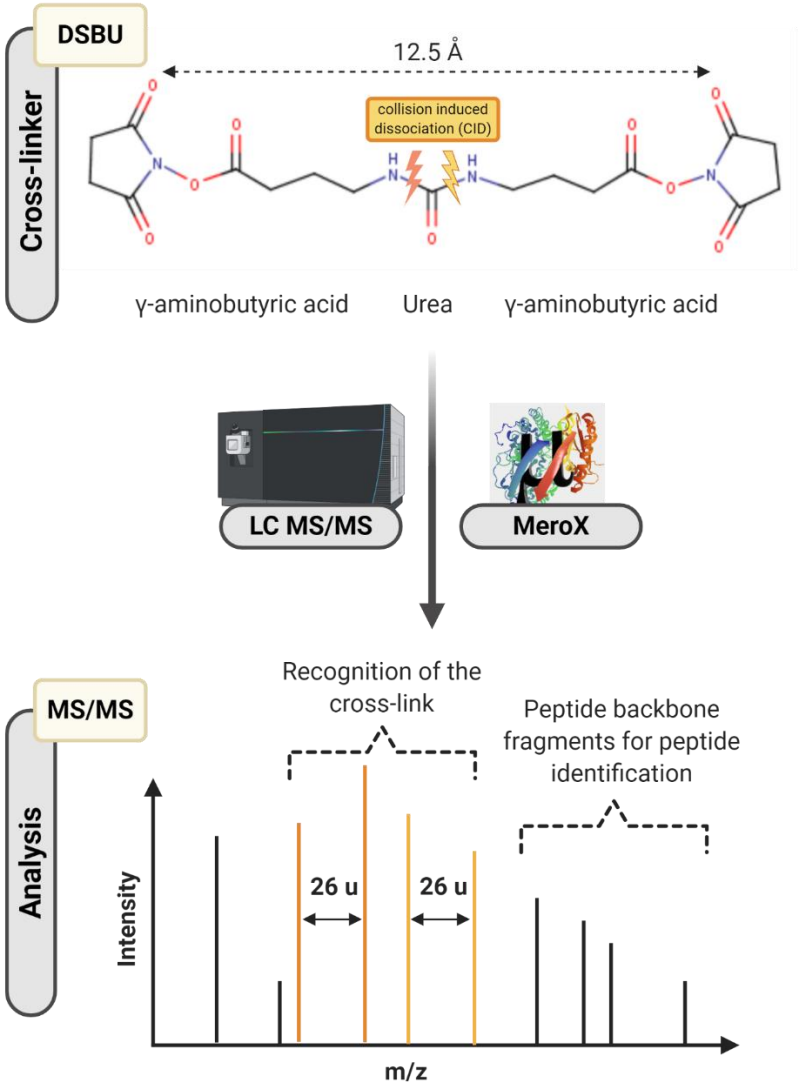



**Figure 33. DSBU structure and cross-linking workflow.** The urea-based cross-linker is 12.5 Å in length, which could be cleaved during the collision-induced dissociation (CID) in 2 γ-aminobutyric acids that fragment in a unique pattern in the gas phase of the LC-MS/MS. This makes it easily identifiable because of its 26-u doublets. The Peptide backbone fragments are essential for peptide identification.

After MS/MS, the cross-linked peptides are reassembled computationally from the identified peptides which are used to provide distance constraints from individual proteins and protein assemblies (Götze *et al.,* 2019). As shown in figure 33, an inter-peptide cross-link will have two distinct doublets of 25.979-u when the cleavage occurs at any of the two NH-CO bonds located at the urea moiety. Therefore, peptides will be modified, with the fragment of the cross-linker in the MS/MS spectra, showing signals from the two doublets.

Obtained primary MS/MS data was subsequently evaluated computer-based using MeroX. However, each data set was inspected manually for quality and unambiguity. The cross-linking sites were manually compared by selecting various combinations and inspecting the spectrum. Notably, the spectrum should contain fragment ions of both peptides, fragment ions from the fragmentation of the cross-linker (DSBU), ion series with high intensities, and a minute number of unassigned peaks (which could be due to co-isolation or neutral losses) (Götze *et al.,* 2019). All fragment ions have unique colours in the detailed result sheet, which further helped in the inspection. Scores are generated by MeroX depending on the presence and intensity of the reporter ions. Additionally, the length of the ion series of the peptide backbone and the number of ion fragments in comparison to the theoretical peptide pairs generated (Yu & Huang 2018).

The list generated of all cross-links had high scores from 110 down to 10 of a total of 246 cross-links. However, a cut-off score was used, and no cross-link was considered below 50. Analysis of the ≥ 300 kDa band revealed 50 intermolecular cross-links (between two different domains) and 37 intramolecular cross-links (within the same domain) (Supplemental table 23). However, analysis of the ≈ 200 kDa band revealed three intermolecular cross-links and 21 intramolecular cross-links (Supplemental table 24). Thus, the identified cross-links are shown as a circos diagram using CX-Ciros (Chait Lab) (Figure 34a and 34b). The diagram shows all inter-and intra- molecular cross-links. Furthermore, the degree of conservation of each residue was included in figure 34 using ConSruf (Ashkenazy *et al.,* 2016). All intermolecular cross-linked residues were manually checked using Cnx1 sequence with identifier: Q39054 to ensure that the correct numbers are correlated to the correct residues. The numbers and domain location for all cross-links were tabulated for modelling (Supplemental table 24). Upon their identification, the structural availability of the residues was documented using PyMOL (Schrödinger, Inc).

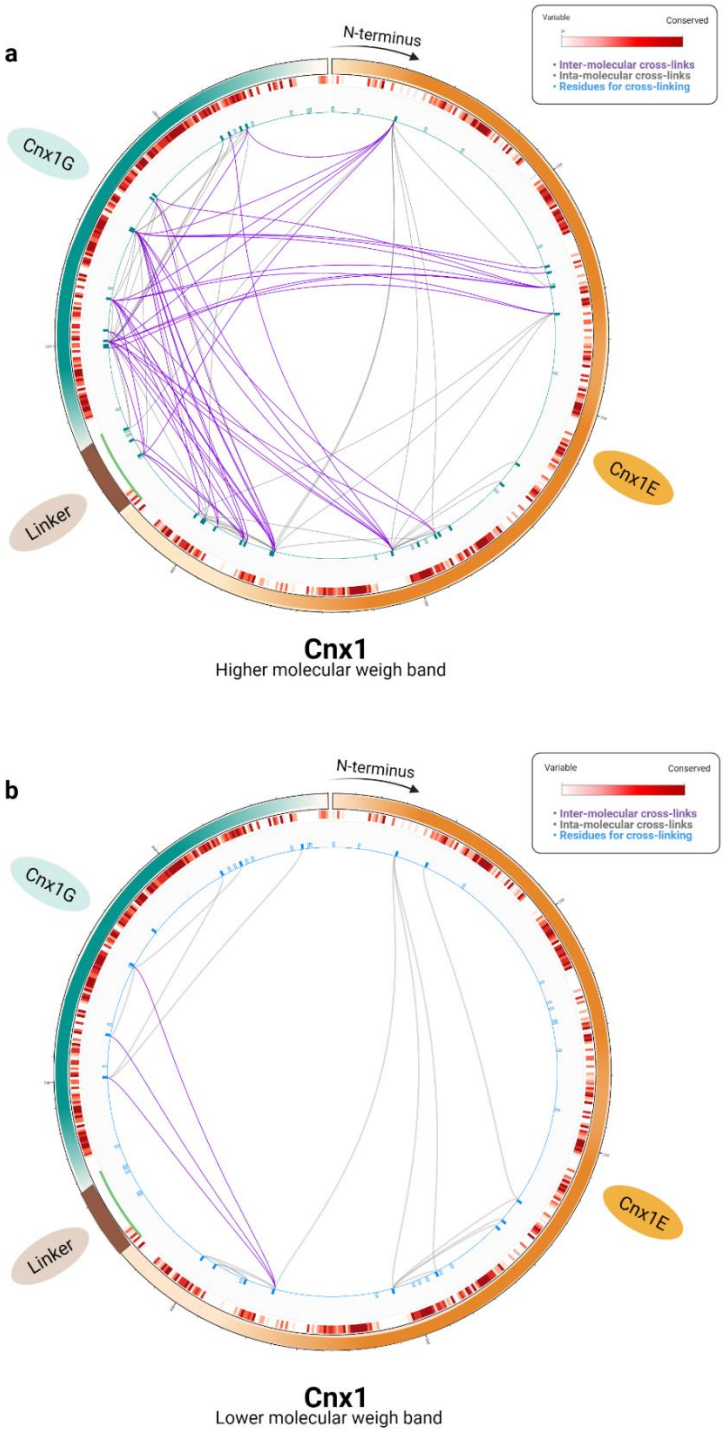



**Figure 34. Circular plots of the cross-links identified for Cnx1 full-length.** The circus diagram shown were generated based on the cross-linking mass spectrometry data obtained for the ≥ 350 kDa complex (a) and the ≈ 200 kDa complex (b), respectively. All intermolecular cross-links were highlighted in purple, and the intramolecular cross-links were highlighted in grey. All lysine residues are highlighted in blue in the diagram, and the degree of conservation of the residues is coloured from highly conserved (red) to non-conserved (white) (Ashkenazy *et al.,* 2016). (a) The circular plot shows a total of 87 cross-links, either intra cross-links between E-domain (bright orange), G-domain (teal), and linker (brown), or inter cross-links within the domain itself. (b) The circular plot shows a total of 21 cross-links, either intra cross-links between E-domain (bright orange), G-domain (teal), and linker (brown), or inter cross-links within the domain itself. The circular plot was generated by CX-Circos (Chait Lab).

## 3.2.4. Modelling Cnx1

Based on the results obtains from XL-MS, the structure of the Cnx1 full-length protein was reconstructed using HADDOCK (Koukos *et al.*, 2020). Therefore, the available crystal structures for Cnx1E PDB: 5G2R (Krausze *et al.,* 2017) and the Cnx1G PDB: 1UUY (Kuper *et al.,* 2004) were used (see 2.2.7.3.). Notably, the crystal structures of Cnx1E and Cnx1G lacks parts of the sequence (Figure 35), including part of the Cnx1E N-terminus, the Cnx1G C-terminus and the domain linker. These gaps of structural knowledge turned out to make modelling of the full-length protein very challenging. Additionally, the structural nature of the domains as homomers provides another challenge to model since the modellers are not equipped to adequately make a distinction between the monomers of a homomeric structure. Therefore, extensive manual validation needed to be done.

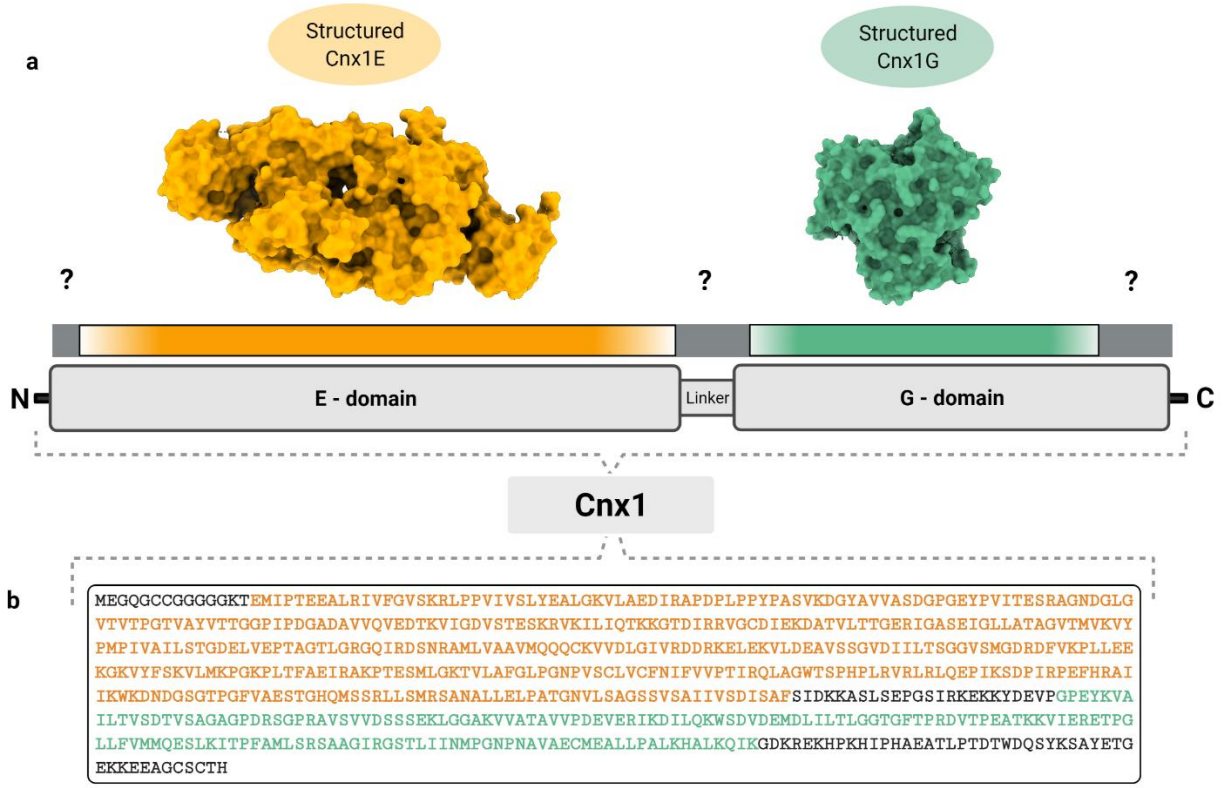



**Figure 35. Structured regions in Cnx1 full-length.** (a) Schematic representation of full-length Cnx1. The structure of Cnx1E (bright orange) (PDB: 5G2R) and Cnx1G (Teal) (PDB: 1UUY) are shown above the schematic domain organization of Cnx1. For parts of the Cnx1 sequence, no structural information is available. These parts are coloured in dark grey, while structurally resolved parts of the sequence are highlighted in yellow and green, respectively. (b) The amino acid sequence of Cnx1 showing the structured E-domain part (bright orange) and structured G- domain part (Teal). Sequence regions for which no structural information is available are coloured in grey (see (a) for comparison).

As described in 2.2.7.3, to prepare TBL files that include the restraints, the cross-linked residues were first marked on the structures. Ambiguously, the crosslinking results from the XL-MS approach showed residue numbers from Cnx1 sequence (with identifier number Q39054) without showing which specific residue on the tertiary structure is linked to which. For instance, residue 158 from the E-domain is cross-linked with residue 521 from the G-domain (on the full-length sequence). As a dimer, the E-domain with residue 158 would be present twice (on the structure). For residue 521, it would appear three times on the trimeric G-domain. This means that the interruption of the cross-links cannot be directly proclaimed on the modeller unless a particular manual visualization is done using structural insights (Figure 36). For this purpose, the modelling was done using all possible realistic combinations. For example, residue 468 in the sequence, which is 7 on the structure of Cnx1G as the first monomer, is also in the other monomer as 171 And third monomer as 335. However, all three cannot be included simultaneously with the same cross-linked residue from Cnx1E.

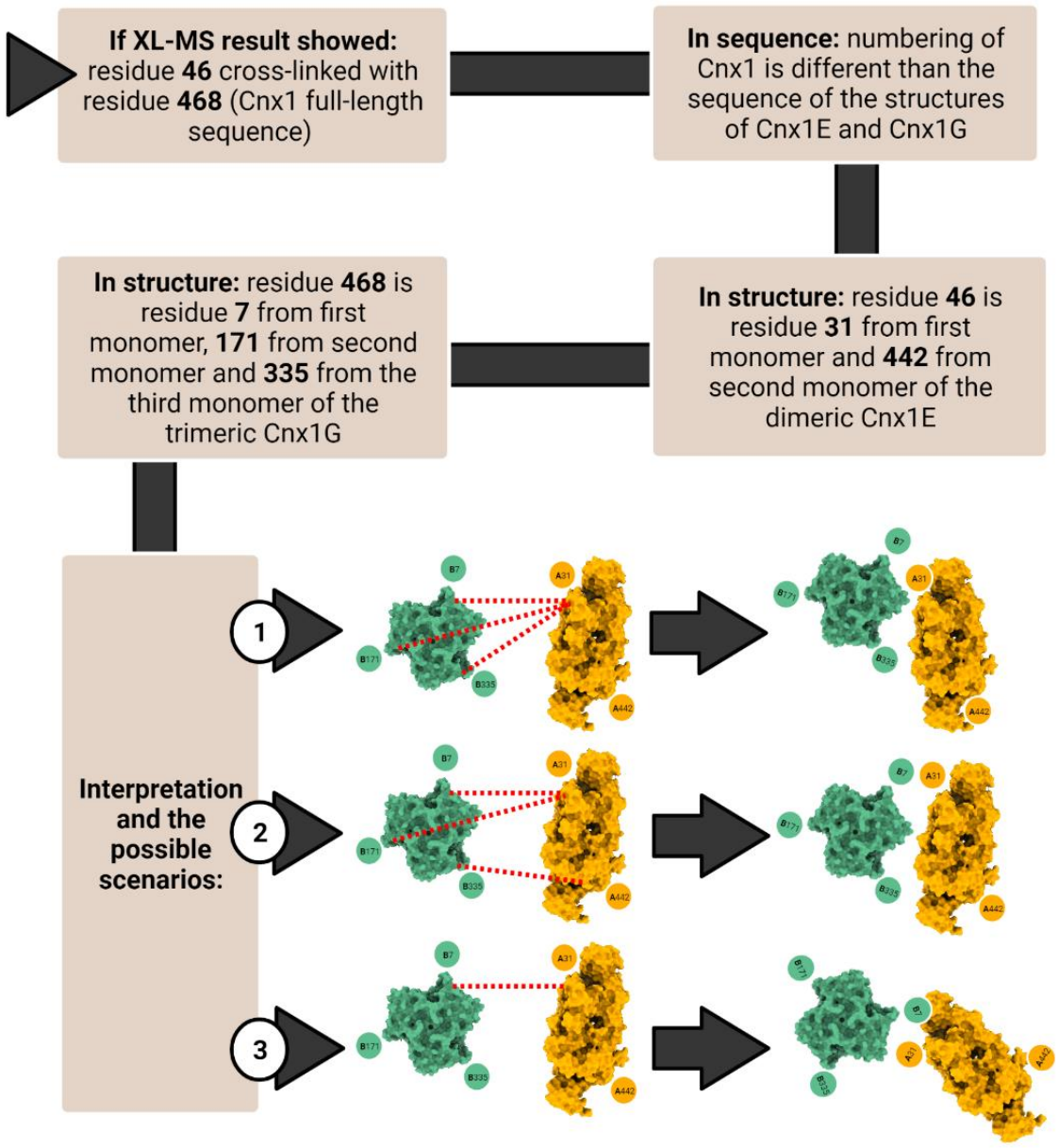



**Figure 36. Restraints modelling interpretation.** For modelling using the obtained restrains, it is essential to differentiate between structure numbering and sequence numbering. The identified cross-linked peptides from the mass spectrometry approaches are given as a number in the full-length sequence without taking into account the 3D structural information (examples are given in the figure). For example, if the results showed that lysine number 46 is inter-cross-linked with lysine 468. This means that residue 468 on the Cnx1 sequence is number 7, 171 and 335 in the trimeric Cnx1G, while 46 is 31 and 442 on the dimeric Cnx1E. All possible scenarios were taken into consideration for modelling. The most realistic scenario was considered to cover the highest number of cross-links.

The cross-links identified for the high molecular weight band are shown in the circos diagram in figure 34 (details in supplemental table 23). Cross-linked amino acid residues

were highlighted in the structures of Cnx1E and Cnx1G using PyMOL (Schrödinger, Inc.), which showed the following:

1) The amino acids identified to be involved in the crosslinking reactions were all traced back to two opposing patches on the Cnx1G surface, 2) Also for Cnx1E, two patches were identified that are located far from each other, 3) Some intra cross-links cannot be directly explained due to irrationality (e.g. intra cross-link that connects a residue with another residue on the other side of the structure, or distance that is not accessible within the structure), 4) some intra cross-links show trimeric or dimeric formation by being crosslinked with the same exact residue (e.g. residue 575 and 556 are cross-linked with residue 575 and 556).

The distance constraints are defined by the distance allowed by the DSBU cross-linker, which has a spacer length of 12.5 Å. This allows the connection between residues with a Cα-Cα range from 4 to 30 Å (Figure 37).

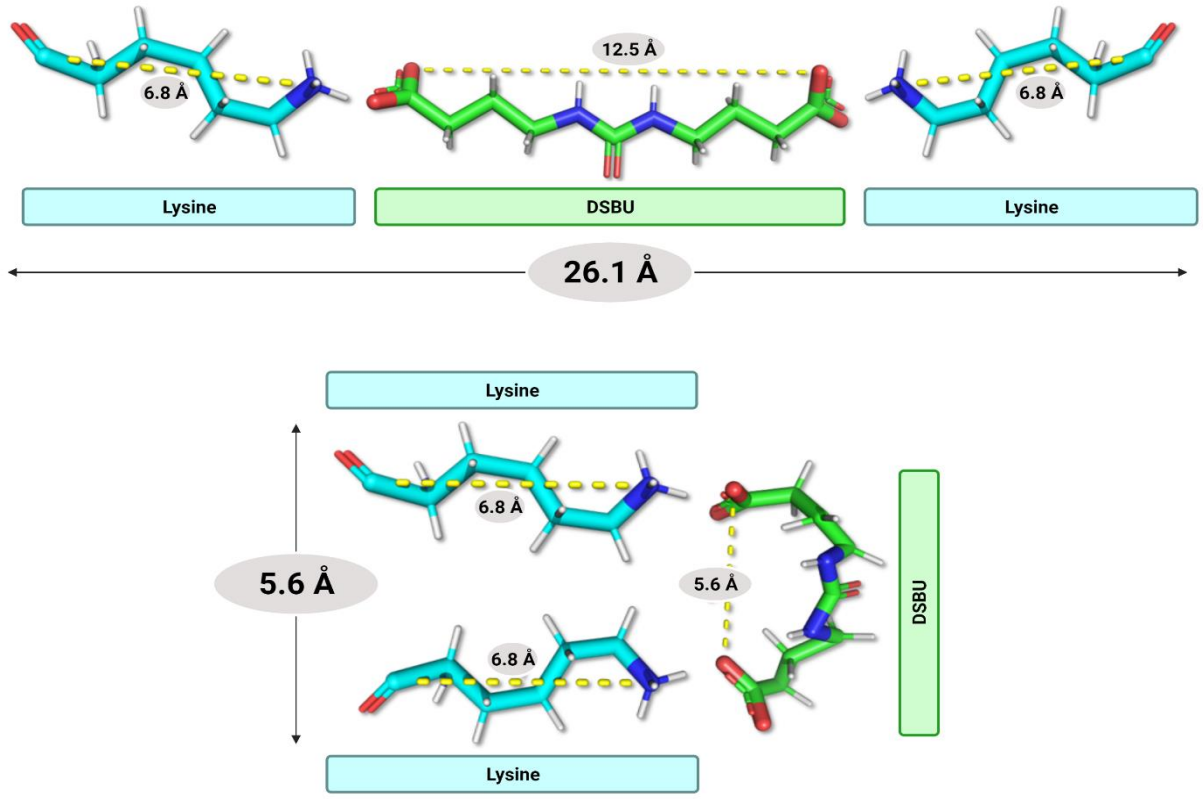



**Figure 37. The range of DSBU cross-linker in cross-linking from 4 to 30 Å Cα-Cα.** DSBU (light green) can cross-link two lysine residues (light cyan). The distance of DSBU is 12.5 Å in fully stretched form. As an example, with the interaction of two lysine residues, each of 6.8 Å the distance, DSBU and the two lysine residues can be cross-linked to 26.1 Å in length Cα-Cα. Another example is in the case of a compact form, DSBU can be in a 5.6 Å range when two parallel lysine residues are interacting closely.

For comparison, the distances were shortened to 4 to 10 Å if necessary to check the modelling results manually. An example of the TBL files prepared can be seen in supplemental table 27.

Due to the different locations of the two sets of cross-links on each side of the protein structure, it was impossible to model all cross-links in one file. These results suggest at least two different orientations of G-domain binding. Therefore, two different TBL files were prepared. Prior to modelling, Cport was used to identify all the active and passive residues involved in the modelling. The active residues are solvent-accessible residues, and passive residues are residues near the active residues identified (Vries *et al.,* 2011). These residues were obtained using the two edited PDB files of Cnx1E and Cnx1G, PDB 5G2R and 1UUY, respectively. The PDB files were edited using coot (rename all chains and numbering if necessary) and PHENIX (to remove all alternative conformations – see 2.2.7.3.) (Emsley *et al.,* 2009; Liebschner *et al.,* 2019).

The results of the active and passive residues of both protein structures are shown in supplemental table 28. The modelling was set to unambiguous, which means all restraints would be enforced. As a comparison, ambiguous modelling was also done, which does not necessarily enforce all restraints in the TBL file uploaded. Afterwards, the model cluster with the lowest energy minimization scores was chosen. The final model was manually checked using selection criteria by checking the distances of all the crosslinks (within the 4 to 30 Å range) using PyMOL as quality control.

The results showed two possible binding sites of the G-domain to the E-domain (Figure 38). As a quality control of the model obtained, I measured the distance from the linker or the distance of the unsolved parts of the E- and G-domain to the solved part of the structures. For example, the C-terminus of Cnx1E (the start of the linker) was measured to the N-terminus of Cnx1G (the end of the linker), the distance was close enough to cover the length of the linker (e.g. ≈ 40 Å). Similarly, with the unsolved C-terminus of Cnx1G, the distance was checked as quality control in the final models to cover an adequate expect distance compensating for the unstructured part.

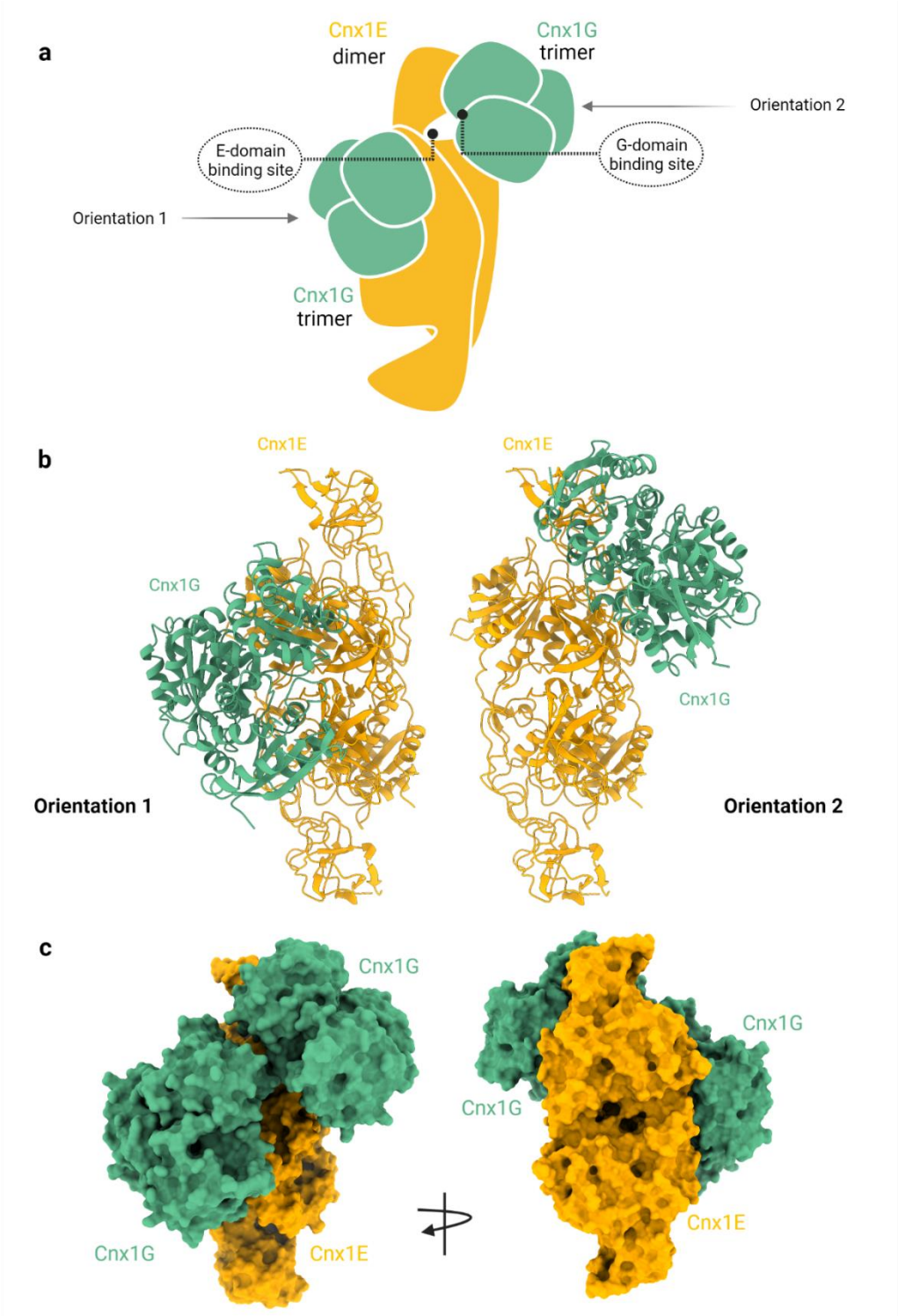



**Figure 38. Modelling the cross-links in Cnx1.** Crystal structures for Cnx1E PDB: 5G2R (Krausze *et al.,* 2017) and the Cnx1G PDB: 1UUY (Kuper *et al.,* 2004) was used as structure input. (a) Schematic representation of restraint modelling. Cnx1E (bright orange) bound to two Cnx1G (teal) molecules from two different orientations. The active site of the E-domain and the G-domain is pointed out with orientation 2 of Cnx1G. (b) Cartoon representation of Cnx1E (bright orange) bound to Cnx1G (teal). Orientation 1 is shown on the left-hand side, while orientation 2 is on the right-hand side. (c) Surface representation of the restraint modelling, including both orientations of Cnx1G bound to Cnx1E.

Further checks for the models were done by checking the intra cross-links (that were not included in modelling). Each intra-domain cross-link was manually measured. However, few of the cross-links showed impossibility. For instance, residue 307 and 158 are cross-linked within the E-domain, however, on the structure, these two residues are 53.1 Å apart (Figure 39). Furthermore, the intra cross-link of residue 501 to 556 within Cnx1G is impossible to be explained since the cross-linker has to go within the structure itself. These unexplained intra-domain cross-links became explainable upon analysis of the oligomerization of the model. As shown in figure 39, an oligomer of the two orientation model with two added Cnx1E molecules explained the riddling intra-cross-links. Residues 307 and 158 that are cross-linked in Cnx1E showed 53.1 Å within the same molecule but 35.2 Å within two different molecules. In Cnx1G, residues 501 and 556 are 38.3 Å apart within the same molecule, however, the distance is 27 Å from the two different molecules.

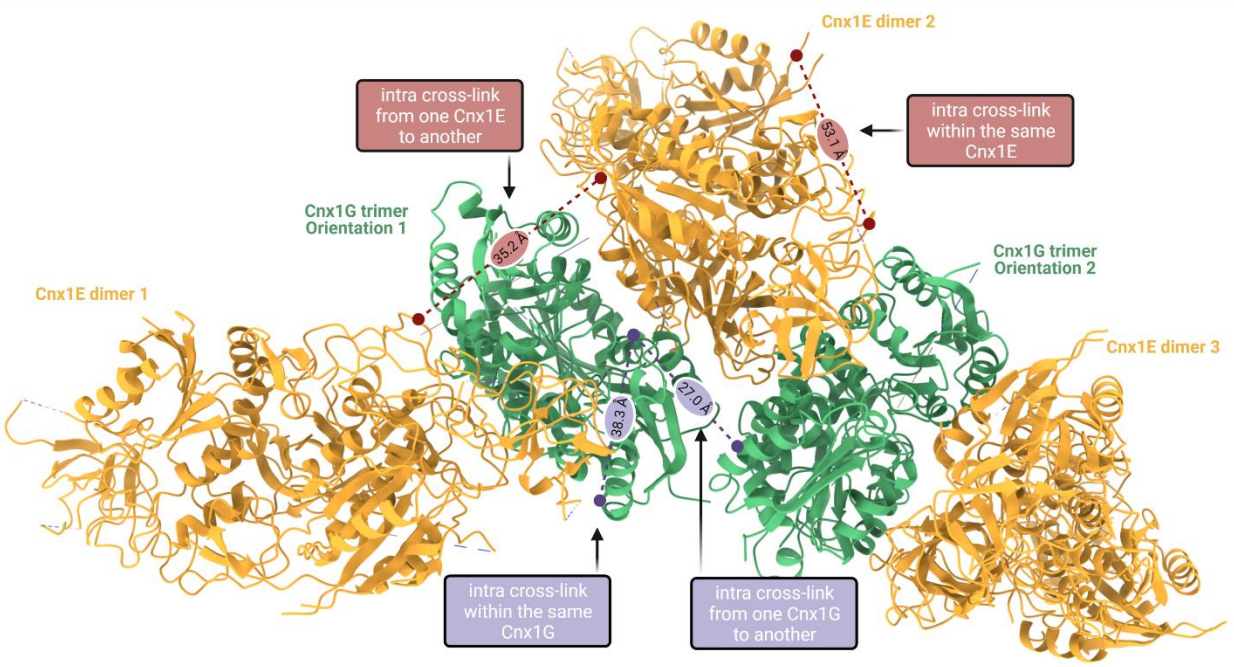



**Figure 39. Assisting the intra cross-links using the oligomerization of the restraint model.** Three molecules of Cnx1E dimer (bright orange) and two Cnx1G (trimers) were used. For Cnx1E dimer 2, both orientation 1 and orientation 2 was included as the model previously shown in figure 38, which is used as the base to assemble the oligomer shown here. The labelled orientation 1 and 2 is based on the two orientations from Cnx1E dimer 2 perspectives. Cnx1E dimer 1 is bound to Cnx1G in the orientation 2 conformation, while, Cnx1E dimer 3 is bound to Cnx1G in the orientation 1 conformation. By considering the Cnx1E-Cnx1G oligomerization states, it was possible to confirm all hitherto non explained cross-link results. Two examples of intra cross-links that agree with the oligomer but not with the structures individually are shown. Residue307 is intra cross-linked with 158 in Cnx1E, which could not be explained assuming the cross-linker to stretch in a distance of 53.1 Å within one Cnx1E molecule, however, 35.2 Å within two different Cnx1E molecules is possible. Another example is residue 501, which is intra cross-linked to 556. The distance within one Cnx1G molecule would be 38.3 Å while 27 Å within two different Cnx1G molecules.

The oligomerization of the resulted model can show an extended super complex. This can be vertically elongated or horizontally extended as a mesh-like complex (as seen in figure 40). Moreover, for Cnx1 full-length oligomer to be formed, the ratio of Cnx1G to Cnx1E must be 1:1.5 (trimer: dimer). If one monomer of Cnx1G is connected to a monomer of Cnx1E, then 3 dimers of Cnx1E is needed with 2 Cnx1G trimers for a sealed assembly unit. As shown in figure 40a, this oligomerization needs at least 2 trimers of Cnx1G and 3 dimers of Cnx1E to form the basic building block, which means 6 monomers of Cnx1G (3x 2) and 6 monomers of Cnx1E (2x 3). The assembly of this building block with another identical one (Figure 40b) can be explained by one intra-crosslink that is found on the E-domain (residue 307 to residue 401 in the E-domain – supplemental table 23). This complex assembly would be a closed unit of 4 Cnx1G and 6 Cnx1E. Using this building block (2 Cnx1G and 3 Cnx1E), different complexes can be formed.

However, *in vitro* experiments with minimal degradation products can still give rise to more fragmented complexes (Figure 41). For instance, the presence of 2 trimeric Cnx1G in the degradation (without the linker or with incomplete linker region), the Cnx1 complex would be composed of 1:1 or 1:2 (trimer: dimer) ratio instead of 1:1.5 when the linker is present. Additionally, one cannot exclude the presence of an indigenous MogA (from the expression of the proteins in bacteria) present in the complex.

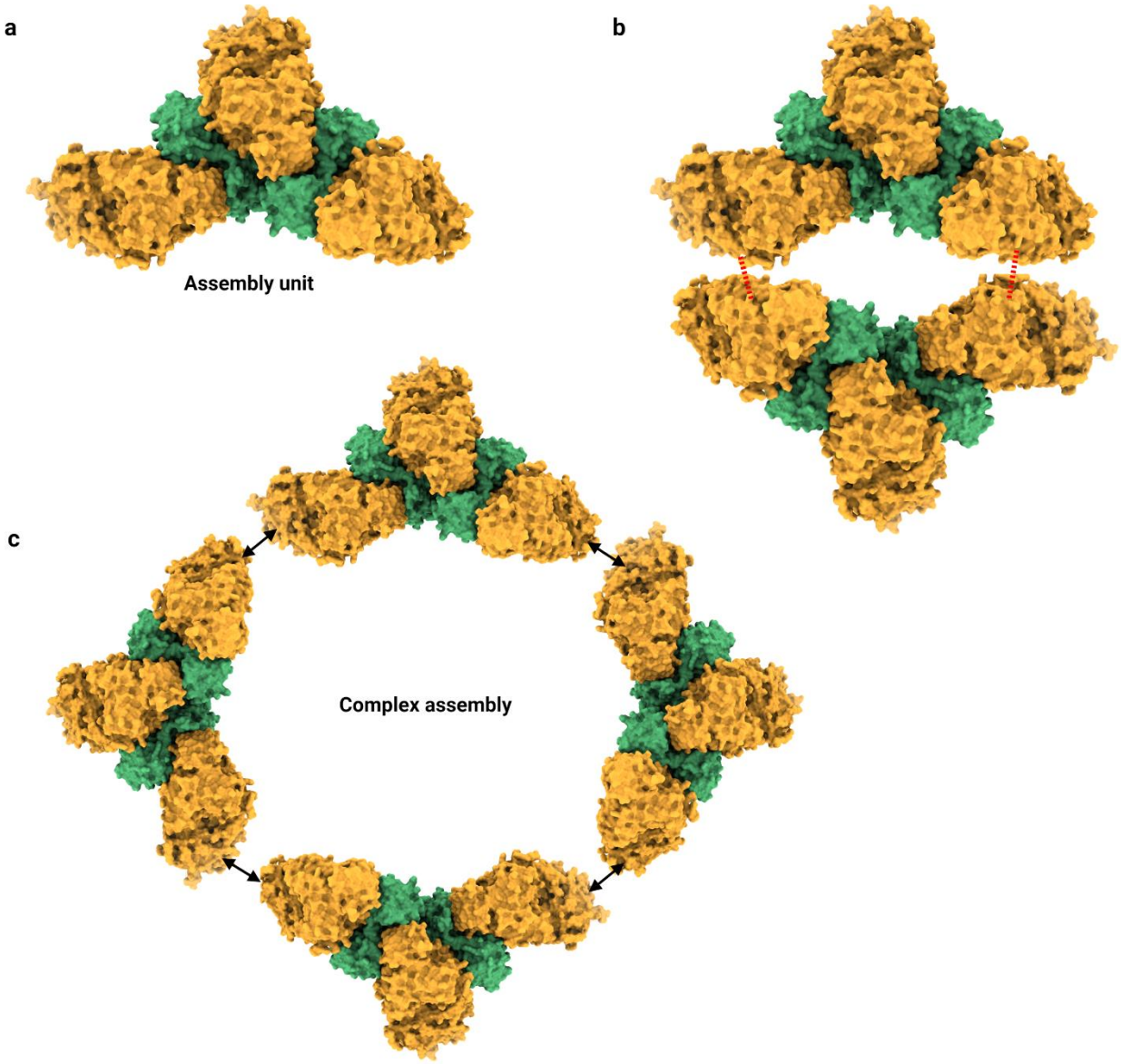



**Figure 40. Oligomerization of full-length Cnx1.** The oligomer shown is based on the final restraint modelling of Cnx1. The basic building block of full-length Cnx1 would be composed of Cnx1E dimer (bright orange) bound to two different Cnx1G (Teal) (orientation 1 and 2). Each Cnx1G trimer would be bound to an additional Cnx1E dimer. (a) The basic building block of the Cnx1 complex including two Cnx1G trimers and 3 Cnx1E dimers. This complex is composed of six full-length Cnx1 molecules. (b) Complex assembly of two building blocks together highlighting one intra-cross-link between two different Cnx1E (dotted red lines). (c) An example of the complex assembly of multiple building blocks to form bigger complexes of full-length Cnx1.

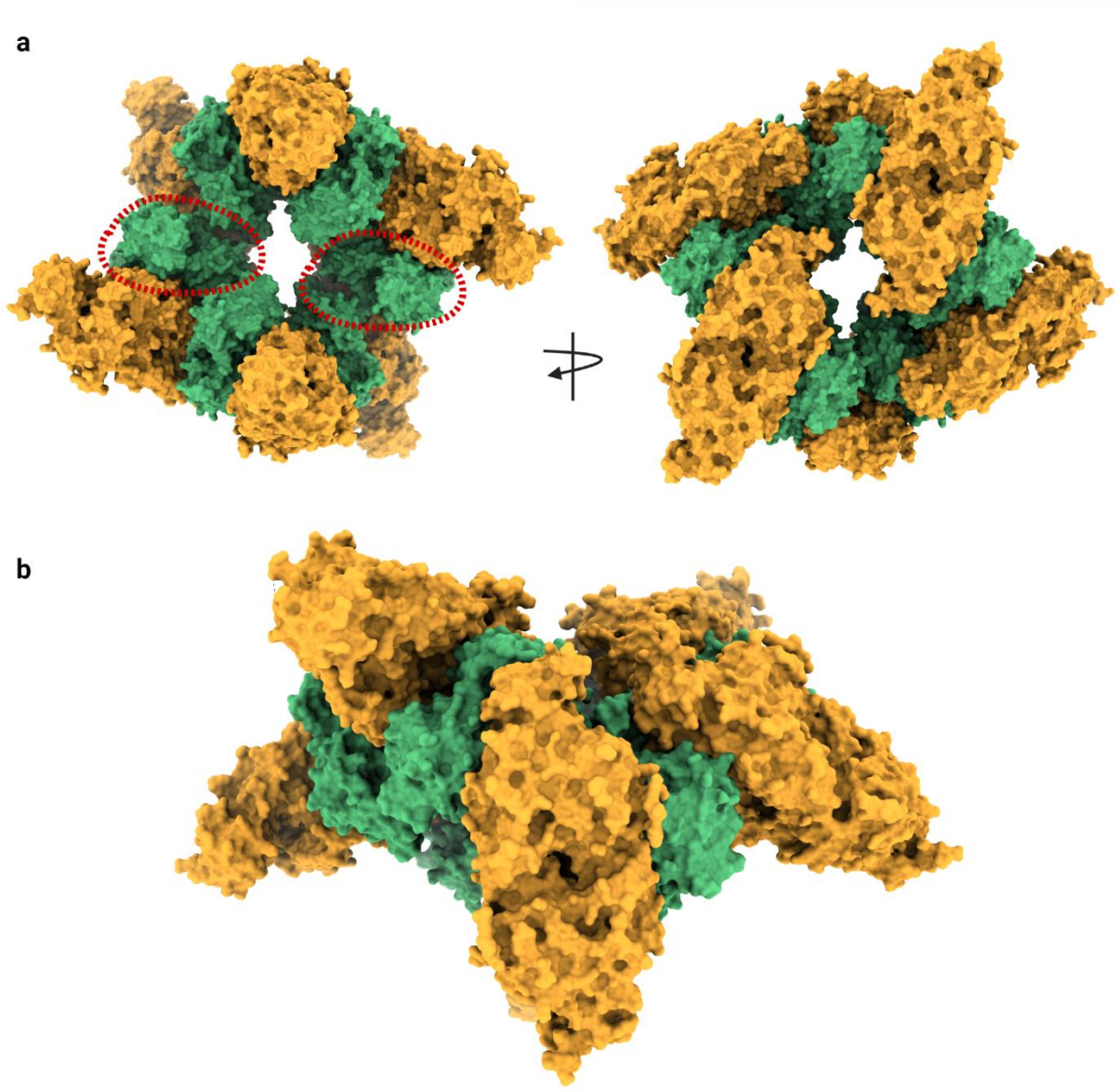



**Figure 41. Cnx1E-Cnx1G complex assembly in the presence of free G-domain.** The oligomer shown is based on the final restraint modelling of Cnx1. Each Cnx1E dimer (bright orange) is bound to two different Cnx1G (Teal) (orientation 1 and 2), which is bound to another Cnx1E (the other orientation). The protein complex assembly shown here consists of six Cnx1E dimers and six Cnx1G trimers. (a) Top and bottom view of the oligomer model highlighting two additional Cnx1G (red dotted circle). (b) side view of the oligomer showing plenty of space for other interacting partners to bind to the oligomer.

Remarkably, the modelling of the XL-MS resembles only a snapshot of what happens in reality. In a dynamic cellular context, there is continuous motion and conformational changes of the proteins (Yu & Huang 2018). Therefore, the two models of the various orientations shown do not necessarily mean the absolute and only existence of the complex

in this form. A slight movement or transition is possible for both domains, and perhaps a transient complex is present or multiple complex forms exist simultaneously in nature as previously shown in other protein complexes (Bullock *et al.,* 2018). The complex formed when Cnx1 takes MPT from the MPT-synthase may be a complex that is different from what it looks like when it adenylates MPT to form Moco-AMP and releases Moco. It is technically impossible to establish fully defined conditions that would allow the elucidation of an MPT-AMP transfer complex.

Moreover, with different interaction partners, the complex may be rearranged. It cannot be excluded that minor re-arrangments that occur would result in a different orientation of E- and G-domain interactions during the different stages of Moco biosynthesis. Considering varying orientations of E- and G-domain interactions, the linker has to allow certain flexibility enabling domain re-arrangements. In gephyrin, it was shown that truncating the linker leads to a decrease in Moco activity while removing the linker leads to depletion (Belaidi & Schwarz 2013). However, it is not clear how a shortened linker affects Cnx1 functionality. The modelling data from the MoeA-MogA complex also resembles the metabolic transformation model of Cnx1.

It is essential to model the Cnx1 linker to understand how Cnx1 would look in a cellular biological context. The structure of Cnx1 full-length was predicted using *ab-initio* modelling by I-TASSER or Robetta (Roy *et al.,* 2010; Simons *et al.,* 1999). I-TASSER or iterative threading assembly refinement is an approach in a hierarchical form in order to predict protein structure. Initially, the template of the known part of Cnx1 (Cnx1E and Cnx1G structures) is considered by a multiple threading approach called LOMETS (Roy *et al.,* 2010). The models are then constructed via iterative assembly simulations. The resulted models of full-length Cnx1 were manually checked using PyMOL (Figure 42). The model that showed no clashes in a protein complex assembly of Cnx1 was taken into consideration. In a second line of work, Robetta was used for comparative modelling and *ab-initio* modelling. Structure prediction is made using a deep learning-based method called TrRosetta or Rosetta CM (Simons *et al.,* 1999). For Rosetta CM, four independent methods are involved, including RaptorX, HHpred, Sparks-X, and Map align (Chivian *et al.,* 2003; Kim *et al.,* 2004). Rosetta AB is used for *ab-initio* modelling, and the resulting models were compared to I-TASSER, which was manually checked with the final HADDOCK-modelled complex of Cnx1 (Figure 42).

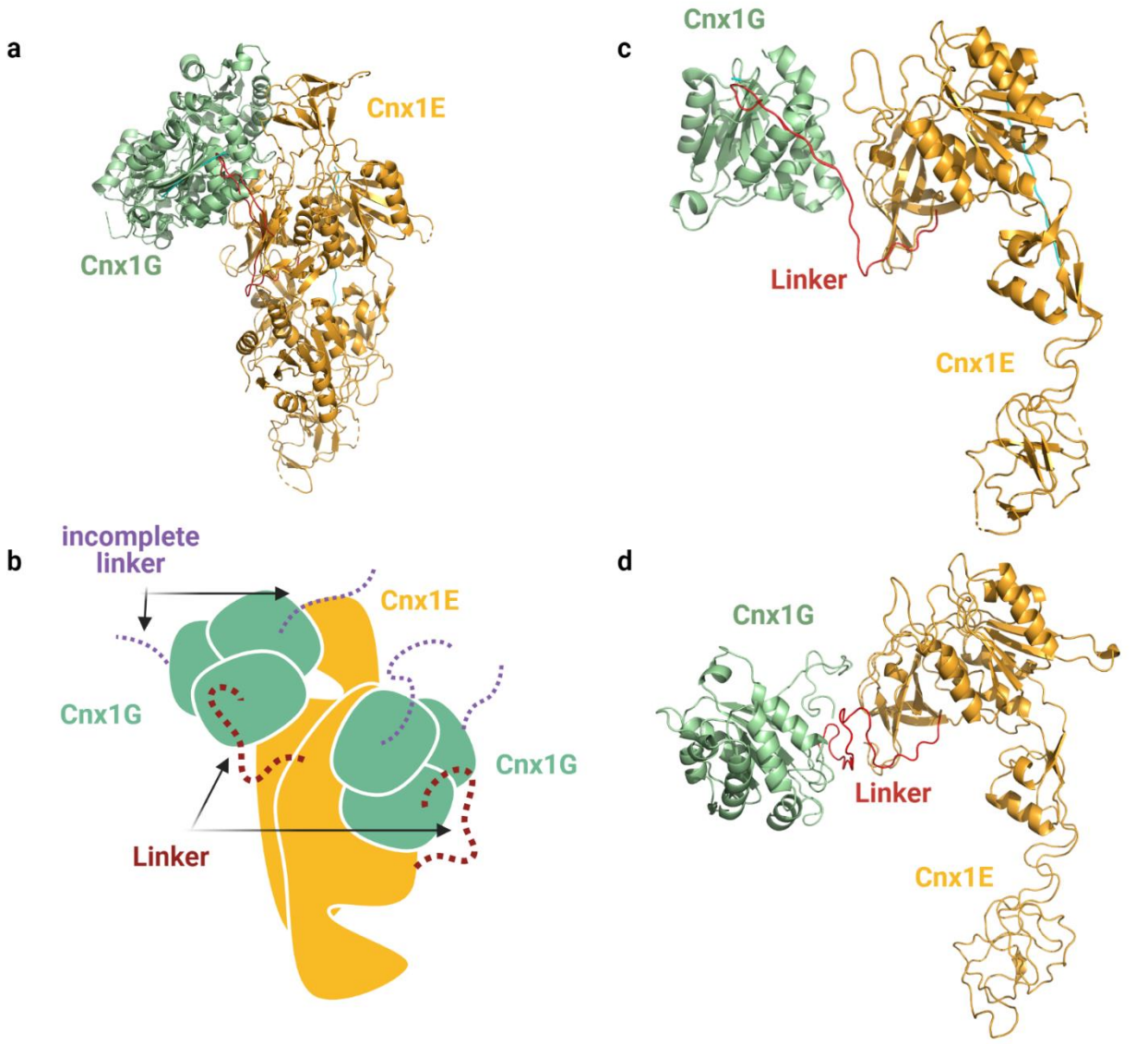



**Figure 42. The *ab initio* modelling of Cnx1 linker.** Modelling was done with I-TASSER (Roy *et al.,* 2010) and Robetta (Simons *et al.,* 1999). The final restraint model was used as a template for the structure part of Cnx1. (a) Orientation 2 (metabolic transfer) of Cnx1G (teal) bound to Cnx1E (bright orange). The linker modelled is highlighted in red (b) Schematic representation of orientation 1 and 2 with the linker in red and incomplete linker in purple. (c) Model of the resulting full-length Cnx1 using I-TASSER. (d) Modelling of the Cnx1 linker using Robetta.

Both models generated from I-TASSER and Robetta were compatible with the oligomerization models. However, Robetta model showed a lower RMSD compared to the model from I-TASSER.

Upon assembly, the E-G domain complex must be fully capable of interacting with both, upstream and downstream enzymes involved in the cellular Mo-metabolism. Considering

the G-E domain complex reported in this work is existent in the cell under conditions of Moco demand, the MPT-synthase must be capable of interacting with it to ensure the supply with the substrate, i.e. MPT. To give an answer to this question, docking of the Cnx1 model generated in this work with MPT synthase complex (PDB 3BII) was done. Due to the lack of a crystal structure of a plant (*A. thaliana*) MPT synthase complex, a template-based PHYRE2 model was predicted based on the structure of the prokaryotic (*E. coli*) MPT-synthase complex was used for the docking approach (see 2.2.7.). Both docking results were almost identical since PHYRE2 probably used PDB: 3BII as a template due to the high homology between both sequences of the MPT synthase in *E. coli* and *A. thaliana*. The Modelling was done with three different Cnx1 models. Firstly, orientation 1 only, secondly, orientation 2 only, and finally, both orientation as a final model (Figure 43). Some of the resulting models were excluded due to clashes with the model (e.g. the unsolved part of Cnx1G would clash with the MPT synthase docking site). Thus placed, the MPT-synthase active site was relatively distant to the G-domain active site as the G-domain active site to the E-domain active site.

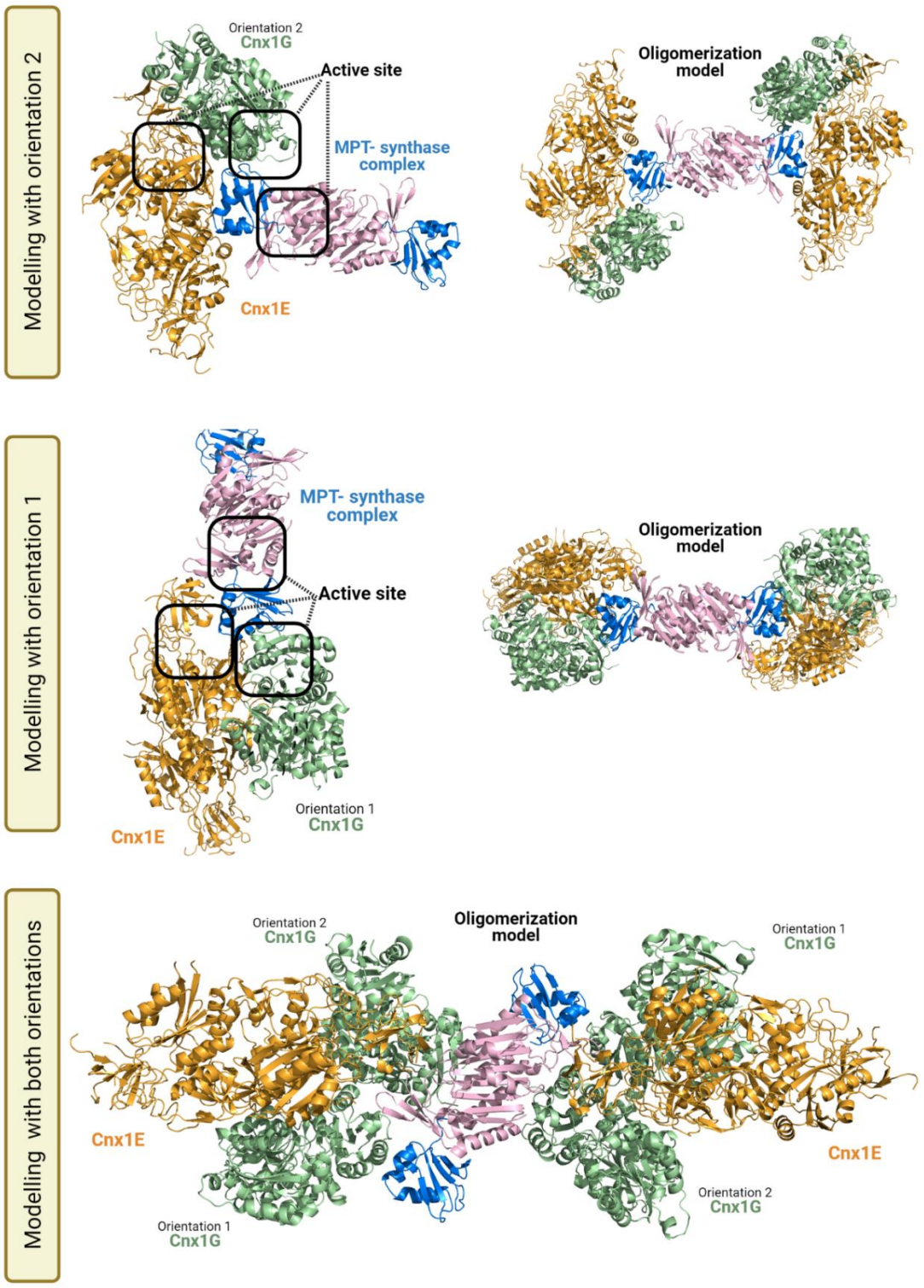



**Figure 43. Docking Cnx1 complex with MPT synthase complex.** Modelling was done using ClusPro using three different approaches. Firstly, docking using orientation 2 (Cnx1E in bright orange and Cnx1G in Teal) as a receptor and MPT-synthase (PDB: 3BII) (Large subunit in light pink and small subunit in marine blue) as a ligand. Secondly, using orientation 1 of the Cnx1 complex as receptor and MPT- synthase as a ligand. Finally, using both orientations as receptor and MPT-synthase as a ligand. The figure highlights the active site of Cnx1E, Cnx1G and the MPT- synthase.

MPT synthase, the MPT-forming complex of the second step in Moco biosynthesis, is composed of two subunits, which shows great structural homology to the eukaryotic ubiquitin. MPT synthase is a heterotetramer of a large dimerized subunit and two small subunits on opposite ends on the large subunit. In close proximity to the negatively charged C-terminus of the large subunit, the active site of MPT synthase lies (Rudolph *et al.,* 2001). With each of the two small subunits, a large positively charged pocket is present, forming the cavity in which MPT is formed. The Cnx1-MPT-synthase model revealed some highly conserved amino acids of opposing charges which were opposing each other (G-domain – MPT-synthase) (Figure 43). It is expected that the MPT synthase interacts with the Cnx1 G-domain to supply the latter with MPT, however, interaction with the E-domain cannot be excluded. Since MPT was shown to bind to Cnx1G with high affinity (Schwarz *et al.*, 1997; Kaufholdt *et al.,* 2016), it can be deduced that the MPT-synthase interact with orientation 1 of Cnx1G bound to Cnx1E, as in this position MPT maybe first adenylated, yielding MPT-AMP. This can be transferred only when Cnx1G interacts in orientation 2 manner with Cnx1E (Figure 44). Importantly, since Cnx1 was shown to be the core of the Moco biosynthesis complex (Kaufholdt *et al.,* 2016), the protein complex assembly shown in figure 40 would be the platform for all MPT synthase binding.

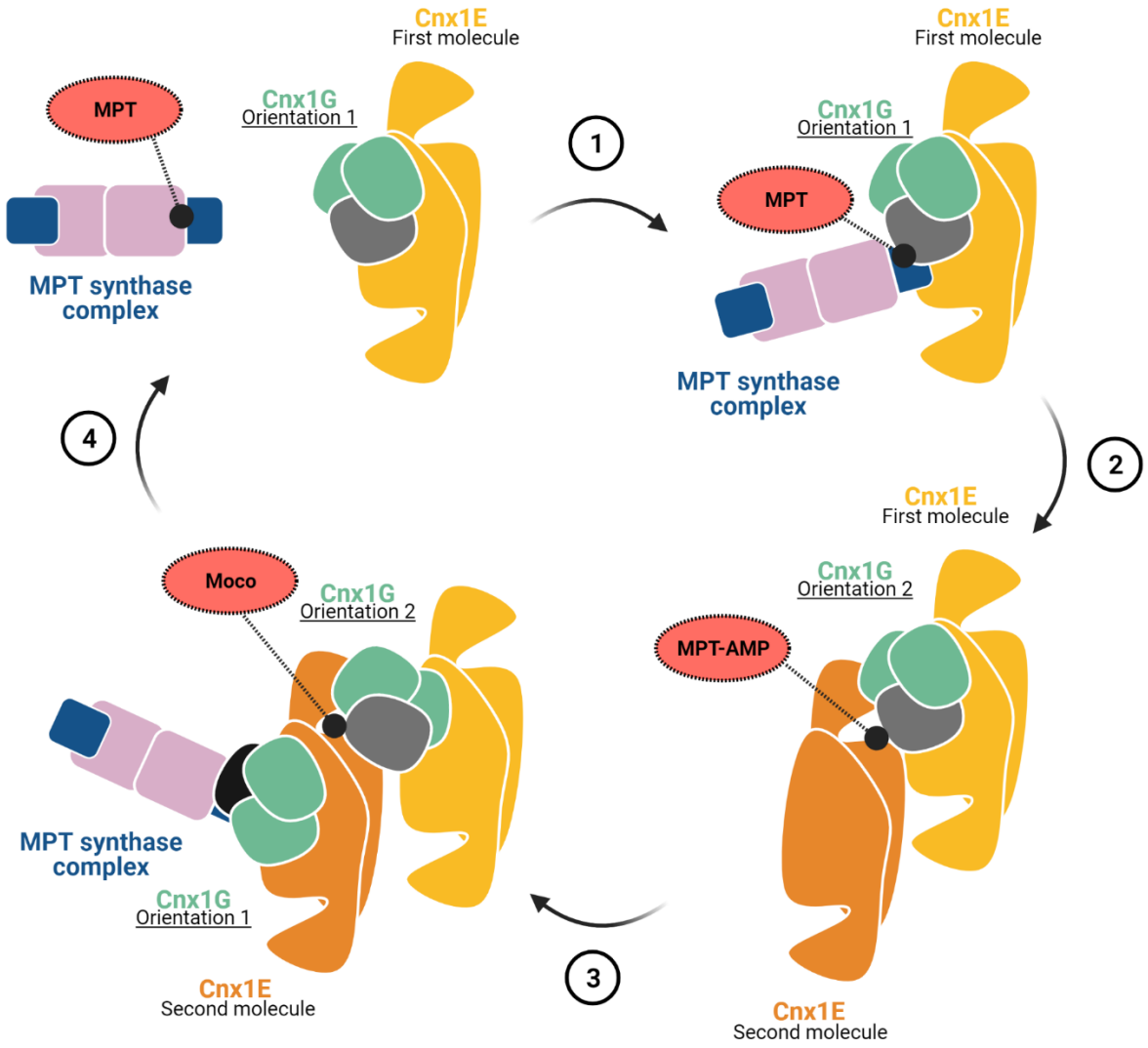



**Figure 44. Schematic illustration of how MPT synthase would interact with Cnx1 to transfer MPT.** The illustration is based on the model shown from docking the Cnx1 complex with MPT synthase complex. MPT synthase complex is represented in light pink (large subunits) and marine blue (small subunits). Initially, MPT synthase (with MPT bound) would interact with one monomer (grey) of the Cnx1G trimer in orientation 1. The MPT is then transferred to the Cnx1G monomer. After adenylation of MPT by Cnx1G (to form MPT-AMP), another Cnx1E dimer would interact with Cnx1G in orientation 2 manner. Subsequently, MPT-AMP is transferred to Cnx1E. Simultaneously, another (or the same) MPT synthase carrying MPT would interact with another monomer of Cnx1G (black). Subsequently, the cycle is repeated for other Cnx1E and Cnx1G molecules.

As one would expect, the active site of Cnx1E is shown to be highly conserved. For interaction with other partners (e.g. Actin), there are few highly conserved and surface-exposed residues in Cnx1E and Cnx1G (Figure 45). These residues are not in the active site directly, in some cases, in close proximity. For instance, the highly conserved set of residues **F**SK**V**L**MKPG**K**P**LT**F** (highly conserved residues highlighted in bold) of surface-exposed phenylalanine (F), methionine (M) and threonine (T). These three residues, when strictly

conserved, showed to be a possible ''hot spot'' for binding of other potential protein interaction partners (Ma *et al.,* 2003). Although the lysine (K) was shown to be important in the structural integrity of Cnx1E (Krausze *et al.* 2017), however, the phenylalanine, methionine and threonine could possibly be a binding site for other Moco-receiving enzymes such as NR (if in close proximity to the binding site of Cnx1E) or the cytoskeleton (not necessarily near the active site of Cnx1E). Importantly, further detailed investigation of the residues needed for complex formation is needed to be addressed as a confirmation to the Cnx1 complex model.

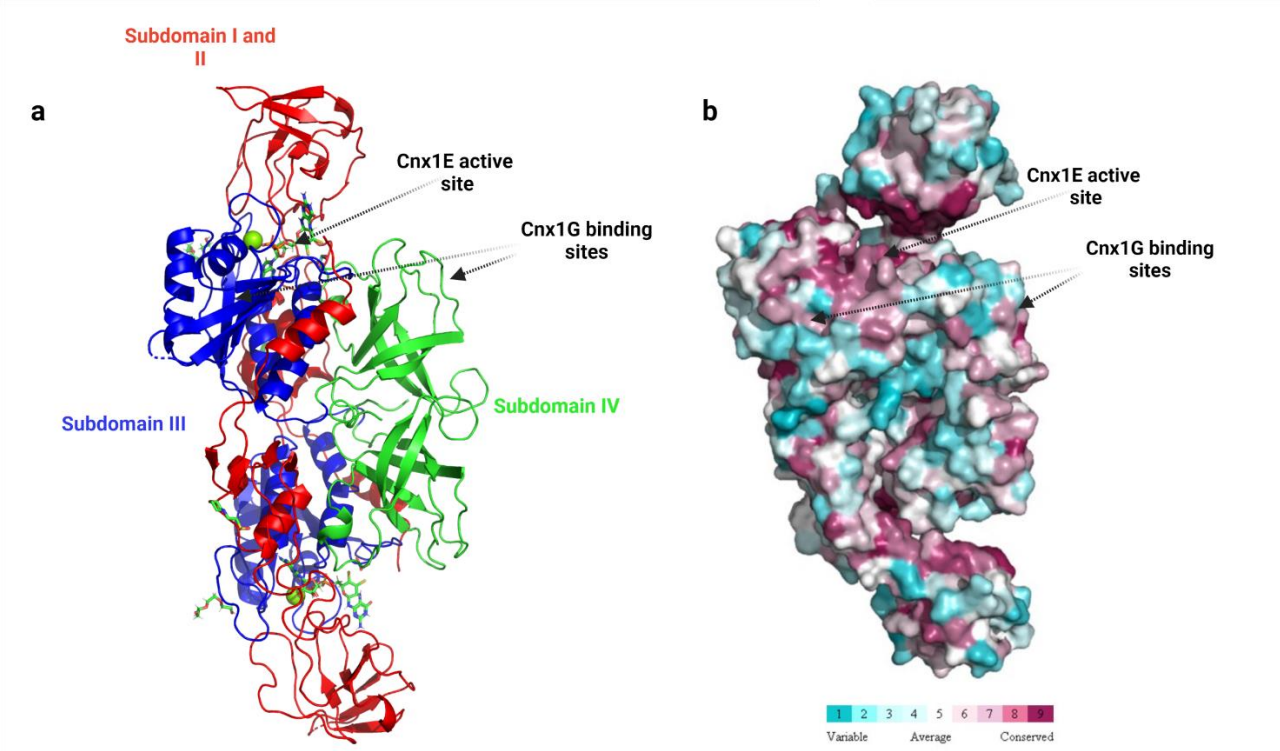



**Figure 45. Subdomain annotation of the E-domain/ conservation of residues.** The structures shown was prepared using Cnx1E PDB: 6Q32. (a) The four Cnx1E subdomains I to IV are shown in red, blue and green for subdomain I-II, III and IV, respectively. Both active site and site of Cnx1G interaction are pointed at with arrows. (b) Cnx1E conservation is done using Consurf (Ashkenazy *et al.,* 2016) with HMMER homolog search algorithm with 1 iteration and E-value cutoff of 0.0001 and UNIREF-90 protein database. The degree of conservation is ranged from variable (cyan), average (white) and conserved (dark magenta).

## 3.2.5. Product-substrate exchange

Based on the model generated in this work, the two orientations in which Cnx1G interact with Cnx1E represent a model of how product-substrate exchange could exist. For orientation 1, the Cnx1G binding site for MPT or MPT-AMP respectively is far apart (≈35 Å) from the Cnx1E active site. Conversely, for orientation 2, both active sites are 15 Å Angstrom apart. Precisely, the distance of the Cnx1G bound petrin moiety (found in MPT-AMP) to the Cnx1E pterin moiety (found in Moco-AMP) is ≈ 16.5 Å (Figure 46a). Therefore, the orientation 2 interaction model may be considered as a metabolite transfer model. However, defined distance excludes the direct transfer of MPT-AMP from the G-domain active site to the E-domain active site. Earlier work showed that Cnx1E is a functional dimer (Krausze *et al.,* 2018), suggesting that Moco-AMP is bound to subdomain III of one dimer and stabilized by the flexible subdomains I+II of the other dimer. Notably, the hitherto presented interaction models did not take into account the all flexible hinge region present in subdomains I+II, which is causal for the flexibility of the subdomain I+II head as documented by large b-factors. Assuming this, I suggest that this flexible head moves apart and hence allows the G-domain and E-domain active sites to align in a productive orientation which allows metabolite transfer between the active sites

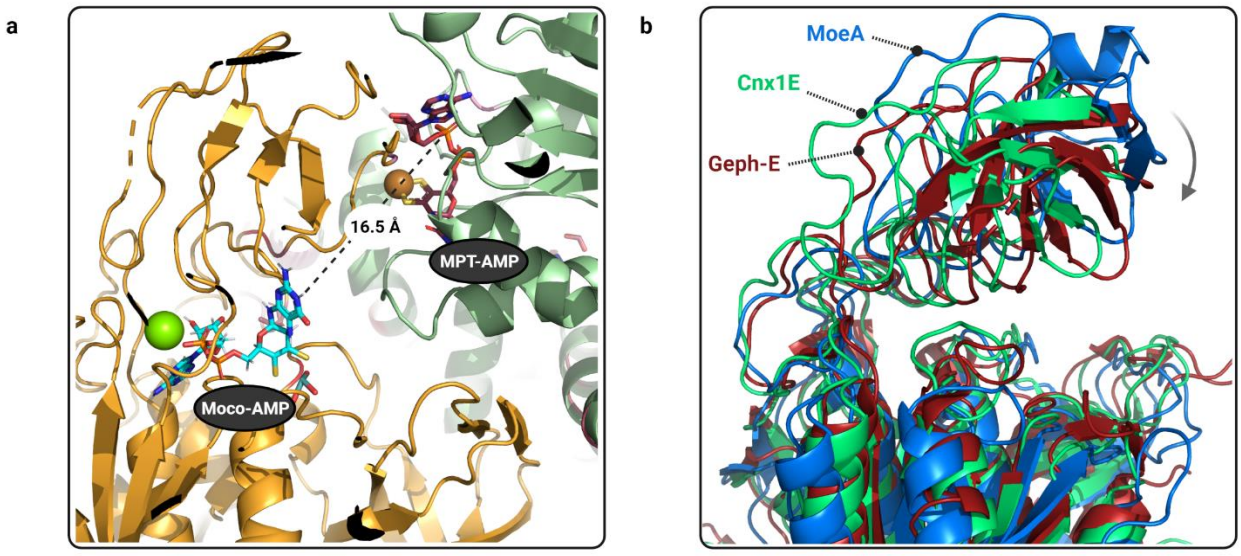



**Figure 46. Close up on the metabolic site of both Cnx1E and Cnx1G.** (a) The model presented in this work of the orientation 2 is shown wherein Cnx1E (bright orange) used with PDB: 6Q32 and Cnx1G (teal) used with PDB: 1UUY. The distance of the Cnx1G bound petrin moiety (found in MPT-AMP) to the Cnx1E pterin moiety (found in Moco-AMP) is ≈ 16.5 Å. (b) Close-up on the hinge region of the E-domain. Superimposition of Geph-E (4TK3), Cnx1E (52GR) and MoeA (this work)

The hinge region of the E-domain protects the active site of Cnx1E (Krausze *et al.,* 2018). The model built does not explain how MPT-AMP is transferred from G-to E-domain, as the distance between the two active sites (G- and E-domain) precludes its directed transfer. A likely explanation was obtained upon analysis of the b-factors and structural characteristics of subdomains I+II (E-domain), which are known to be highly flexible. As shown in figure 46b MoeA (structure solved in this work), the most opened conformation of the hinge region in the different E- domain structures are known. In my model, I suggest that a minor movement of the latter (≈ 3 Å) is required to allow MPT-AMP transfer from G- to E-domain. I cannot exclude that this movement is triggered by the presence of G-domain bound MPT-AMP and/or the simultaneous presence of molybdate.

Therefore, for a successful transfer of MPT-AMP (product of Cnx1G) to Cnx1E, firstly, the hinge region opens for Cnx1G (in orientation 2) to come in close proximity to the active site of Cnx1E. With or without the help of the Cnx1 linker, the MPT-AMP is transferred to the active site cavity of Cnx1E, where Mo is inserted. A schematic illustration of the possible transfer is shown in figure 47.

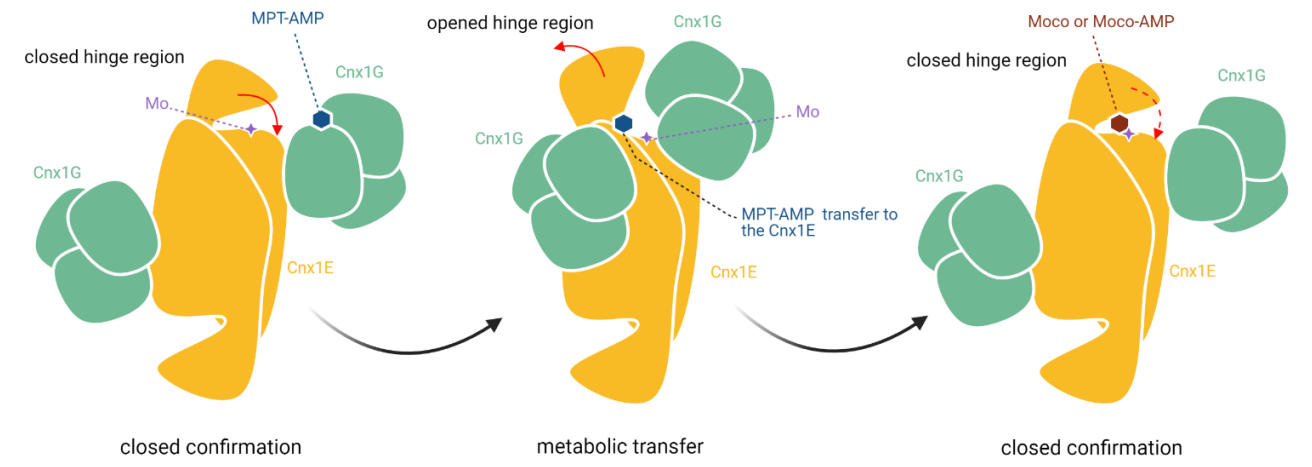



**Figure 47. Schematic representation of how product-substrate channelling propagate.** The two orientation presented in this work is schematically represented of two Cnx1G trimers (teal) bound to one Cnx1E dimer (bright orange). Three-step schematic diagram shows that firstly the closed conformation of the hinge region of Cnx1E with Mo bound and MPT-AMP-bound Cnx1G of orientation 2 interacting with Cnx1E (After receiving MPT from MPT- synthase and adenylation). In the second step (metabolic transfer step), the hinge region is opened, and further movement of Cnx1G from orientation 2 towards the active sites of Cnx1E. MPT-AMP is then transferred to Cnx1E. The third step is the insertion of Mo to form Moco-AMP/ Moco and the closing of the hinge region to protect the product until delivery to user enzymes or further processing.

Notably, the cross-links used for modelling support the proposed metabolic transfer model, as the estimated movements of subdomain I+II would be within the range the cross-linker covers. However, the available docking software did not allow it to define the hinge region as such, which excluded precise modelling of the metabolic transfer complex.

Native MS data of MoeA-MogA showed that a complex is formed in a free environment of Moco metabolites. Although this was previously shown to be the case of plants, however, for Cnx1 as a platform for the Moco biosynthesis complex to be formed, the assembly is needed regardless of the metabolic transfer. Since all cross-linking in Cnx1 was in TP1000 (eukaryotic Moco-biosynthesis containing strain – see 2.1.10.), only one orientation of the Cnx1 model may be present when lacking Moco metabolites, and the other is present only when the metabolites are present (Figure 48). These two orientations of Cnx1G bound to Cnx1E could be existing based on the presence of ligands and the presence of other interaction partners, such as Actin or NR. In order to give the first experimental based insight into the underlying principles of the Mo-insertase E- and G-domain interaction I initially carried out a structural proteomics approach. This approach revealed that the interaction of the prokaryotic and eukaryotic E- and G-domains involves essentially the same parts of the proteins. Notably, for the prokaryotic E- and G-domain, a single interaction model was supported by the evaluated MS-data, while for eukaryotes, a second interaction model became evident. This second interaction model has no apparent function for metabolic transfer between the G and E domain.

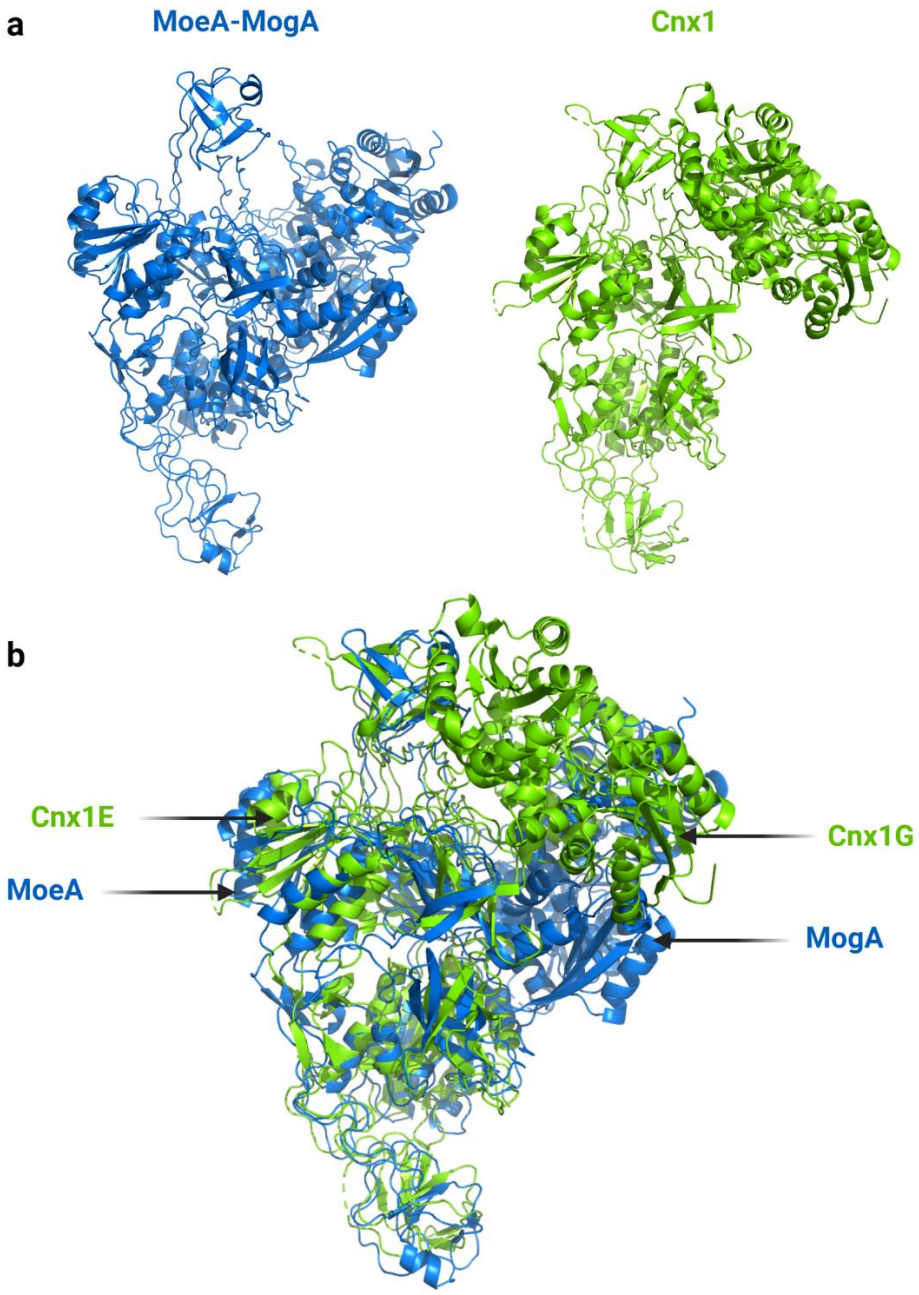



**Figure 48. The final models of Orientation 2 in plants and bacteria.** Models generated from XL-MS –based modelling. Structures used of PDB: 52RS (Cnx1E) and 1UUY (Cnx1G) with MoeA-MogA structures solved in this work (blue). (a) Side by side comparison of MoeA-MogA complex next to Cnx1 complex (resembling orientation 2). (b) Superimposition of both Cnx1 complex (orientation 2) and MoeA-MogA complex.

The fully defined in vitro conditions I applied ensured that no Moco-metabolite was present in the analyzed approaches, hence documenting that even in the absence of Moco

metabolites occurs (at least in vitro). However, I cannot exclude that complex formation is enhanced in the presence of either MPT (MPT-Synthase) or MPT-AMP (Cnx1G). Earlier reports (Magalon *et al.*, 2002) reported that this may well be the case for prokaryotes. My subsequent work (3.3) analyzed both orientations of Cnx1G to investigate if they have a putative role as they may be essentially required for the directed interaction with the MPT-synthase.

## 3.2.6. Negative staining of Cnx1

Modelling the Cnx1E-G complex (Figure 38 and 40) using XL-MS restraints for the computer-based docking approach resulted in two different Cnx1G orientations to bind to Cnx1E. However, since this approach essentially depends on the structural information available for Cnx1E and Cnx1G it naturally has limits as, e.g. the linker region and flexible parts of the proteins may not be considered for model building. To circumvent this possible occurring problem, I next employed negative staining-EM to solve the complex assembly of E-and G domain using recombinant and highly pure Cnx1 full-length protein. The proteins from the Cnx1 purified proteins (as described in 2.2.2.2) was centrifuged prior to negative staining to avoid the possible aggregation formation, which was tested (see supplemental figure 70). From the ≈ 200 electron micrographs that were recorded, particles were manually selected and subsequently classified (Figure 49).

For the negative staining EM approach, all fractions obtained from the gel filtration purification step for initial negative staining. Doing so revealed fractions A4 and A5 possessed minor heterogeneity compared to the other samples. The subsequently carried out particle 2D classification showed inverted ''L'' and ''T''-shaped particles that resemble the two orientation model generated from HADDOCK (see 3.2.4.) (Figure 49c and 49d). However, further Cryo-EM data collection would be required to do 3D modelling.

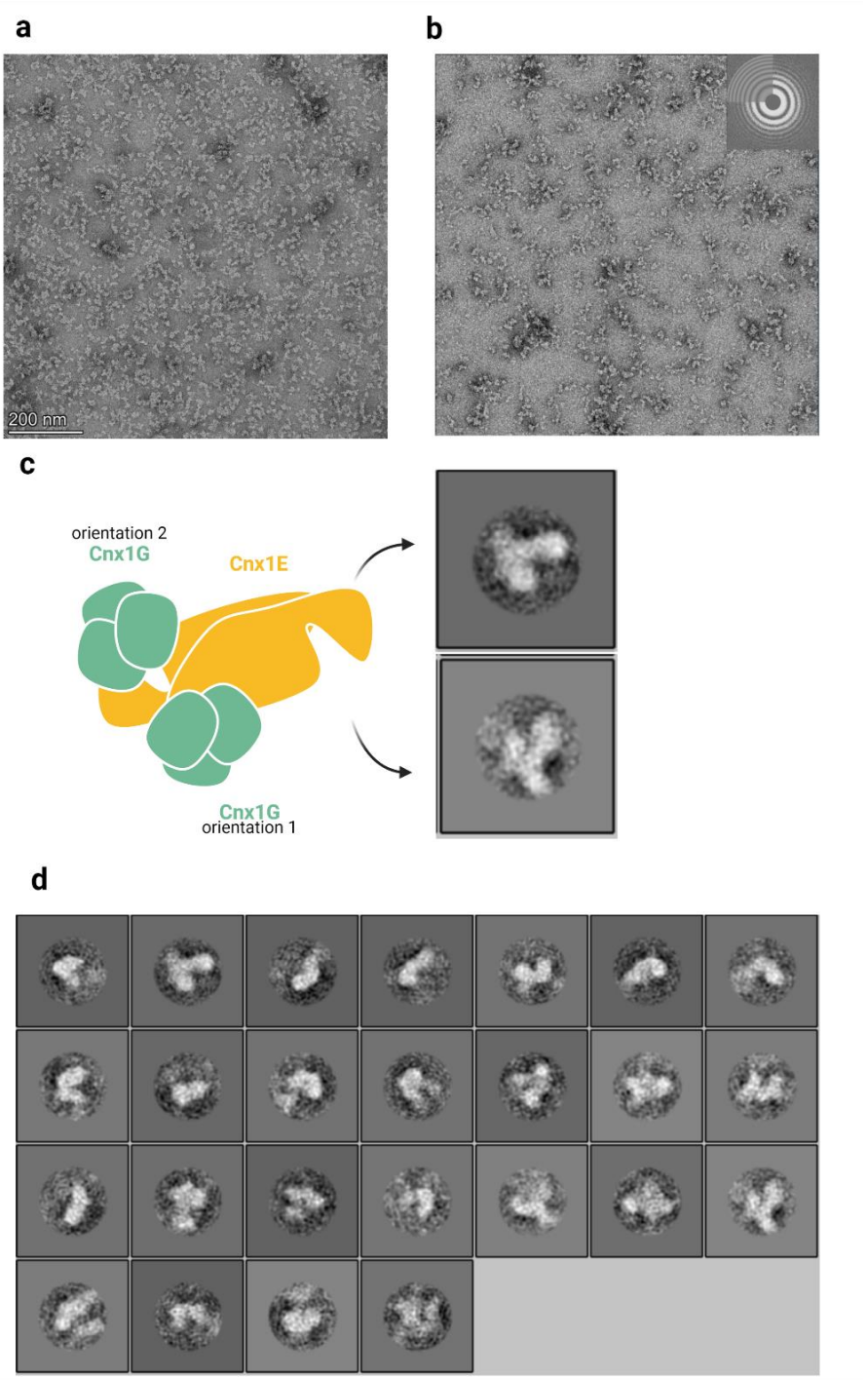



**Figure 49. Negative staining of Cnx1 complex.** All samples were cross-linked in 100 mM HEPES, 300 mM NaCl and 1mM EDTA and particle sized of 120 to 200 Å, which became visible upon negative staining. (a) Negative staining of the Cnx1 complex using recombinant protein with a concentration of 100 µg/ml. (b) Negative staining of the Cnx1 complex using recombinant protein with a concentration of 200 µl/ml. From the data are shown in (b) 129 micrographs were collected, and approximately 45000 particles have been used for classification. (c) Schematic representation of the Cnx1 complex (two orientations) showing two Cnx1G trimers (teal) and one Cnx1E dimer (bright orange). Next to the schematic representation of the Cnx1 complex, representative ''T'' (upper figure) or ''L''- shaped (lower figure) particles obtained from particle classification are shown. (d) Results from the 2D classification of the Cnx1 complexes shown in (b). The following parameters were applied: Boxsize = 200, mask diameter = 300 A. In total, from 45077 particles selected (from b), 25 classes were generated. (Negative staining EM results obtained and processed with the help of T. Beck and P. Wendler).

### 3.2.7 Summary

In summary, a Cnx1 full-length complex was identified and characterized in this work by employing different techniques, including XL-MS, negative staining and computational modelling. Interestingly, the XL-MS-based modelling suggests two Cnx1G binding sites on a single Cnx1E dimer. These results were found to be consistent with results coming from the negative staining EM approach. Thus, any model bias could be excluded.

Side by side comparison with the prokaryotic E-/G-domain interaction allowed it to attribute functions to the different orientations in which Cnx1G interacts with Cnx1E. The orientation 1 model was suggested to serve for the complex assembly and may be required for the upstream interaction of the Mo-insertase complex with the MPT-synthase, while the orientation 2 model was suggested to be required for metabolite transfer between G- and E-domain.

Within the metabolite transfer models (MoeA-MogA and Cnx1E-Cnx1G orientation 2), the E- and G-domain active sites face each other, but the distance between them excludes a direct occurring metabolite transfer. Notably, the Cnx1E / MoeA active sites are shielded by a flexible subdomain of the protein, which prevents the close contact of the two active sites in the computer-based model. Assuming a minor re-arrangement of this subdomain would allow both active sites to come in close proximity towards each other, which is the prerequisite for the direct metabolite transfer between both active sites.

## 3.3. Essential residues for Cnx1 full-length complex

### 3.3.1. Overview

The models generated from the previous approaches shown above gave hints on what exact residues are essential for the complex formation and essential residues required in the substrate-product transfer. To further validate results from structural proteomics / the preliminary Cryo-EM data, I analyzed the Cnx1E and Cnx1G protein surface concerning conserved residues putatively involved in the E-G domain interaction. As expected, the Cnx1E/G active site residues (LIamas *et al.,* 2004; Krausze *et al.,* 2018) are highly conserved. However, next to these, it becomes evident that also on the protein surface, non-active site residues are conserved in both E- and G-domain.

In order to confirm the Cnx1E-G interaction models, I went on and carefully analyzed the interaction interfaces of both interaction models: Cnx1E-G orientation 1 and 2. Doing so revealed numerous strictly or highly conserved residues. Manual inspection of the interaction interfaces identified likely candidates for a site-directed mutagenesis approach, which was subsequently carried out. In the following section, I will describe the identification of E- and G-domain surface-exposed conserved residues that are likely involved in complex formation.

### 3.3.2. Selection of residues for mutation

The selection of the residues was done by initially identifying the residues in the interface of Cnx1G and Cnx1E from the generated model in figure 38. This was done using proteins, interfaces, structures and assemblies or PISA (EMBL-EBI), which results in acquiring structural properties of the quaternary structure assembly as well as their interface (Krissinel & Henrick, 2007). The result shows information such as multimeric state, symmetry number, space group, free energy, accessible or buried surface area, residues, and salt bridges. Fundamentally, oppositely charged residues that are sufficiently close to each other can form a salt bridges bond. For example, the anionic carboxylate group of glutamic (E) or aspartic

acid (D) can form bonds with the cationic ammonium group from lysine (K) or the guanidinium of arginine (R) (Bosshard *et al.,* 2004).

Notably, the control and base for all mutations was a double mutant Cnx1E construct with S269A, D274A mutations (Probst *et al.,* 2021). This double mutant showed high saturations, upon analysis with the HPLC, compared to wild type Cnx1E, which had very low saturations. To see the impact of the complex mutations, high saturations are required in order to see significant alterations. For positive control, D548A mutant was generated, which lies in the active site of Cnx1G (LIamas *et al.,* 2004). This mutant serves as a standard for a substantial MPT-AMP/Moco-AMP reduction. Each of the Cnx1E-G interaction models (Figure 38) was analyzed independently. A summary of the workflow I applied to select residues for mutagenesis is shown in figure 50a.

For orientation 1, 42 residues were identified in the interface of Cnx1E-G, 438 surface-exposed residues, and solvent-accessible area of the interface of 1312.5 Å out of 19195.8 Å in total, 17 possible hydrogen bonds, and seven salt bridges between Cnx1E and Cnx1G. For orientation 2, 53 residues were identified in the interface of Cnx1E-G, 441 surface-exposed residues, and solvent-accessible area of the interface of 1814.7 Å out of 19591.8 Å in total, ten possible hydrogen bonds, and four salt bridges between Cnx1E and Cnx1G. After identifying the interface residues and all possible salt bridges formed, an alignment was done to mine for conserved residues. Using plants, metazoa, fungi, and bacteria, an alignment of all G- and E- domains was done using one class from every phylum in the taxonomy rank. All protein sequences were aligned using cluster omega (EMBL-EBI) while keeping Cnx1E or Cnx1G, Nit-9E or Nit-9G, and MoeA or MogA as a reference. In a second selection round, the identified candidate residues were checked manually for their possible contribution to the E-G domain complex formation.

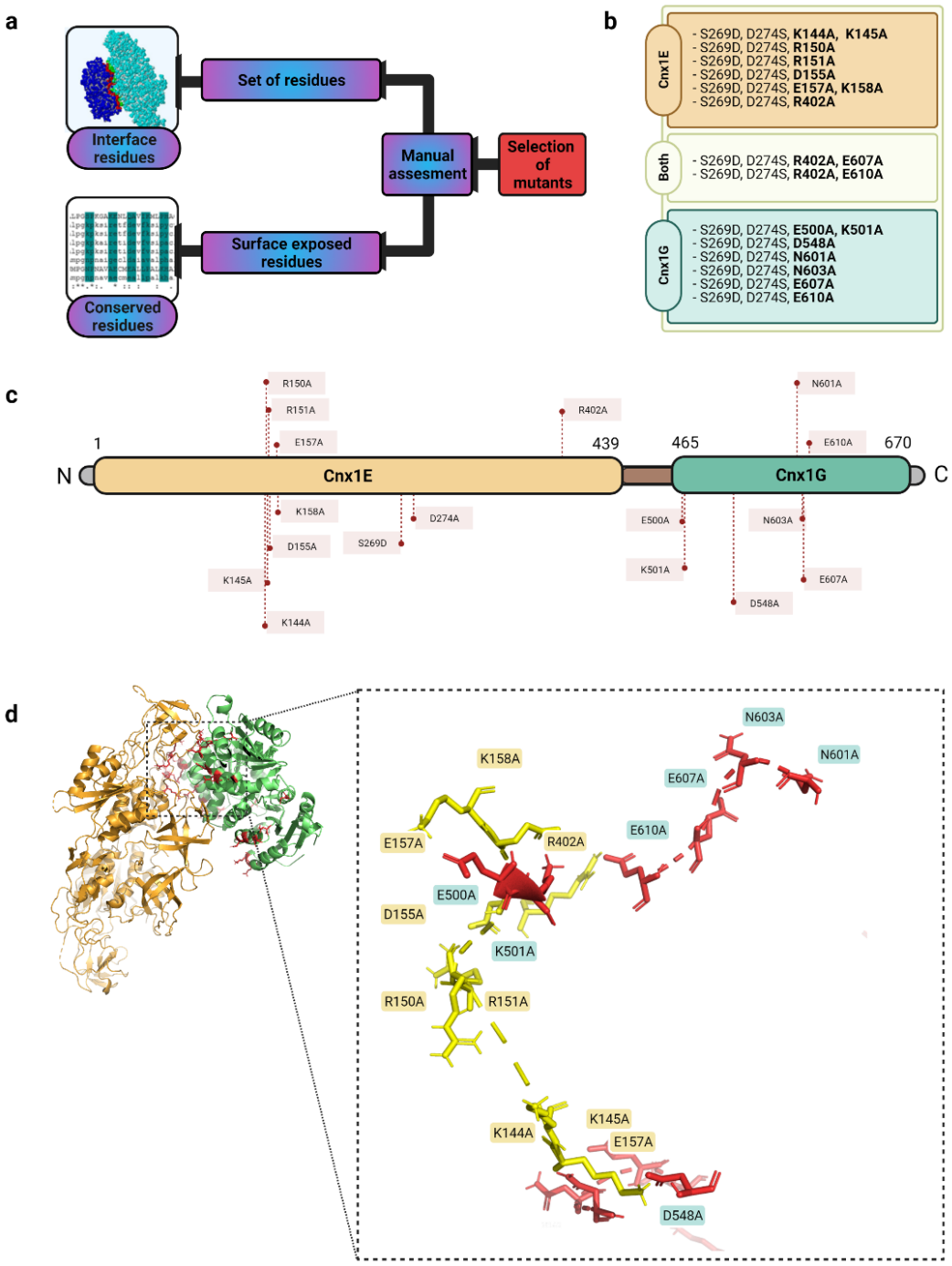



**Figure 50. Summary of the mutants selected for mutagenesis in the metabolic transfer orientation 2.** (a) The workflow describes how candidate residues for structure-guided mutagenesis have been selected. Upon identification of the E-G domain interfaces, conserved residues in Cnx1E and G were identified. These were subsequently used for a manual assessment which comprised structural, functional and biochemical information available for Cnx1E and G. (b) All mutants generated for orientation 2 in the Cnx1 model complex. S269D D274AS is used as a control, six mutants are generated for Cnx1E (pale orange), two combined Cnx1E and Cnx1G mutants, and six Cnx1G mutants were generated. (c) Schematic representation of Cnx1 full-length, all mutants in Cnx1E (pale orange) and Cnx1G (teal) are numbered referring to the Cnx1 (protein identifier Q39054). (d) Orientation 2 of the modelled Cnx1 complex showing Cnx1E (bright orange) bound to Cnx1G (pale green). All residues mutated are highlighted in sticks on the structure. Residues mutated from Cnx1E are shown in yellow and Cnx1G mutants in red.

For orientation 1, the seven salt bridges identified from the G-domain were used as the candidates for mutagenesis. The residues were E518, K521, D522, R547, K556 and E559. One of the six residues to be mutated showed a capability of salt bridge formation in two different monomers of the G-domain. Thus, six residues mutated should result in seven salt-bridge disruptions (Figure 51). A gene was synthesized (BioCat) with the six mutated residues. On the other hand, for orientation 2, highly conserved and/or potential salt bridge forming residues in Cnx1E were K144, K145, R150A, R151, D155, E157, K158 and R402, while Cnx1G included E500, K501, D548, N601, N603, E607 and E610 (Figure 50b, 50c and 50d).

The nomenclature used is based on Cnx1 full-length sequence with identifier number: Q39054. All residues of orientation 2 were exchanged to alanine. Alanine was the amino acid residue of choice since it is a non-bulky, short amino acid that tends to form alpha helices and can exist in beta sheets but can also act as a truncated side chain back to the Cβ (Bosshard *et al.,* 2004). However, one of the disadvantages of exchanging multiple alanine residues in a protein is that it leads to a decrease in the solubility of the proteins (Hou *et al.,* 2018). That is the reason that the genes synthesized for orientation 1 had serine exchanges rather than alanine.

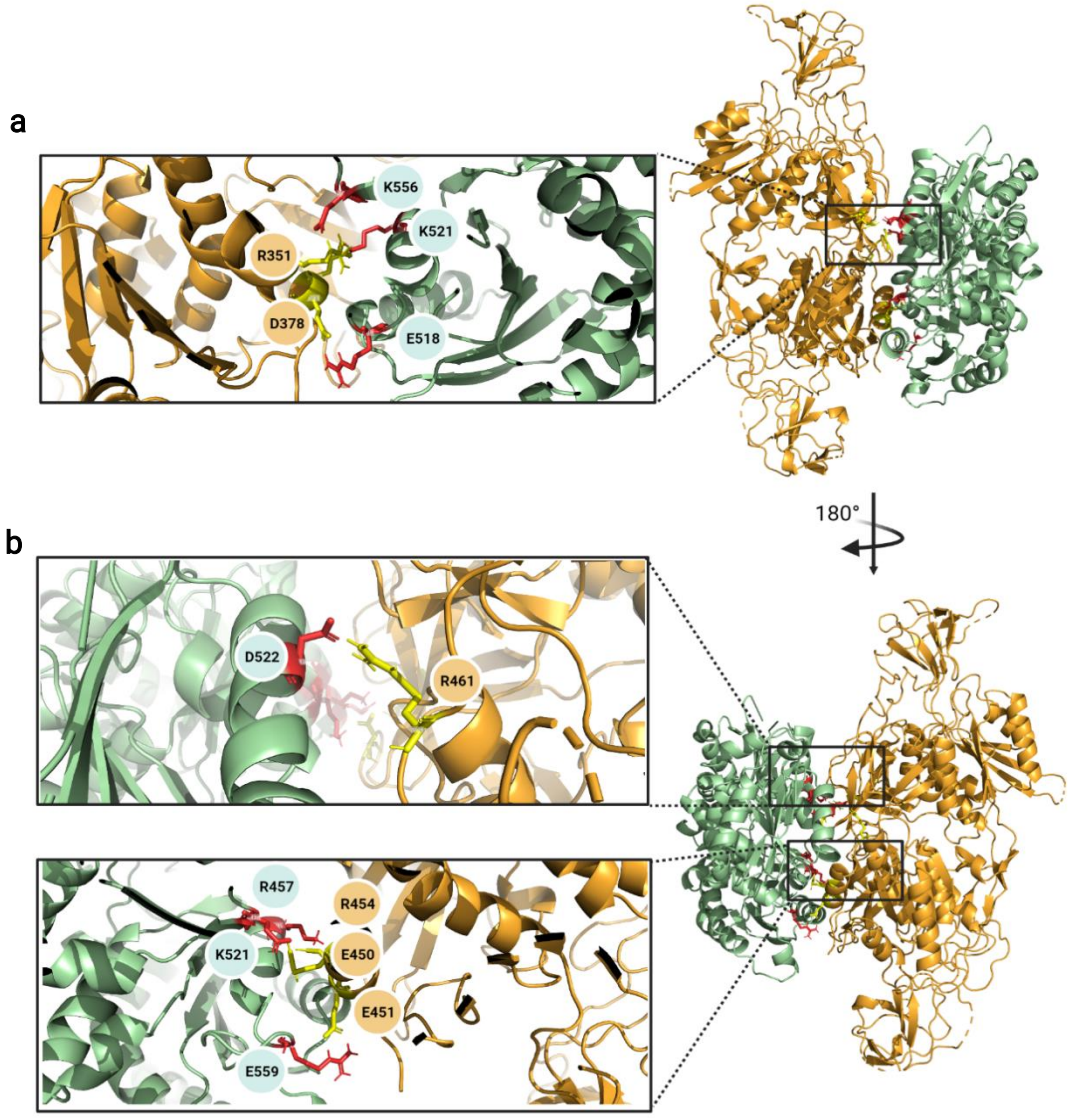



**Figure 51. Summary of the orientation 1 residues targeted by structure-guided mutagenesis.** Orientation 1 of the model generated in this work is shown in a cartoon representation using PDB: 5G2R for Cnx1E (bright orange) and PDB: 1UUY for Cnx1G (teal). Cnx1E residues targeted by mutagenesis are shown as yellow sticks, while Cnx1G residues targeted by mutagenesis are shown as red sticks. The six residues shown for Cnx1G (including E559, K521, R457, D522, K556 and E518) are responsible for a possible 7 salt bridges, with K521 responsible for two different salt bridges on two different Cnx1G monomers with two different residues on Cnx1E. (a) Front view Cnx1G residues K521, K556 and E518, are shown (green). The Cnx1E interacting residues are R351 and D378, respectively. (b) Back view highlighting D522 (Cnx1G) with R461 (Cnx1E), and R457, K521 and E559 (Cnx1E) with R454, E450 and E451 (Cnx1G). All mutants are numbered using the Cnx1 full-length (Q399054) numbering.

## 3.3.3.Site-directed mutagenesis results and discussion

In the following, all variants created (Figure 50 and 51) were expressed, purified, and analyzed with respect to protein purity and Moco metabolite loading. The results obtained for orientation 1 and 2 exchange variants are summarized (Figure 52).

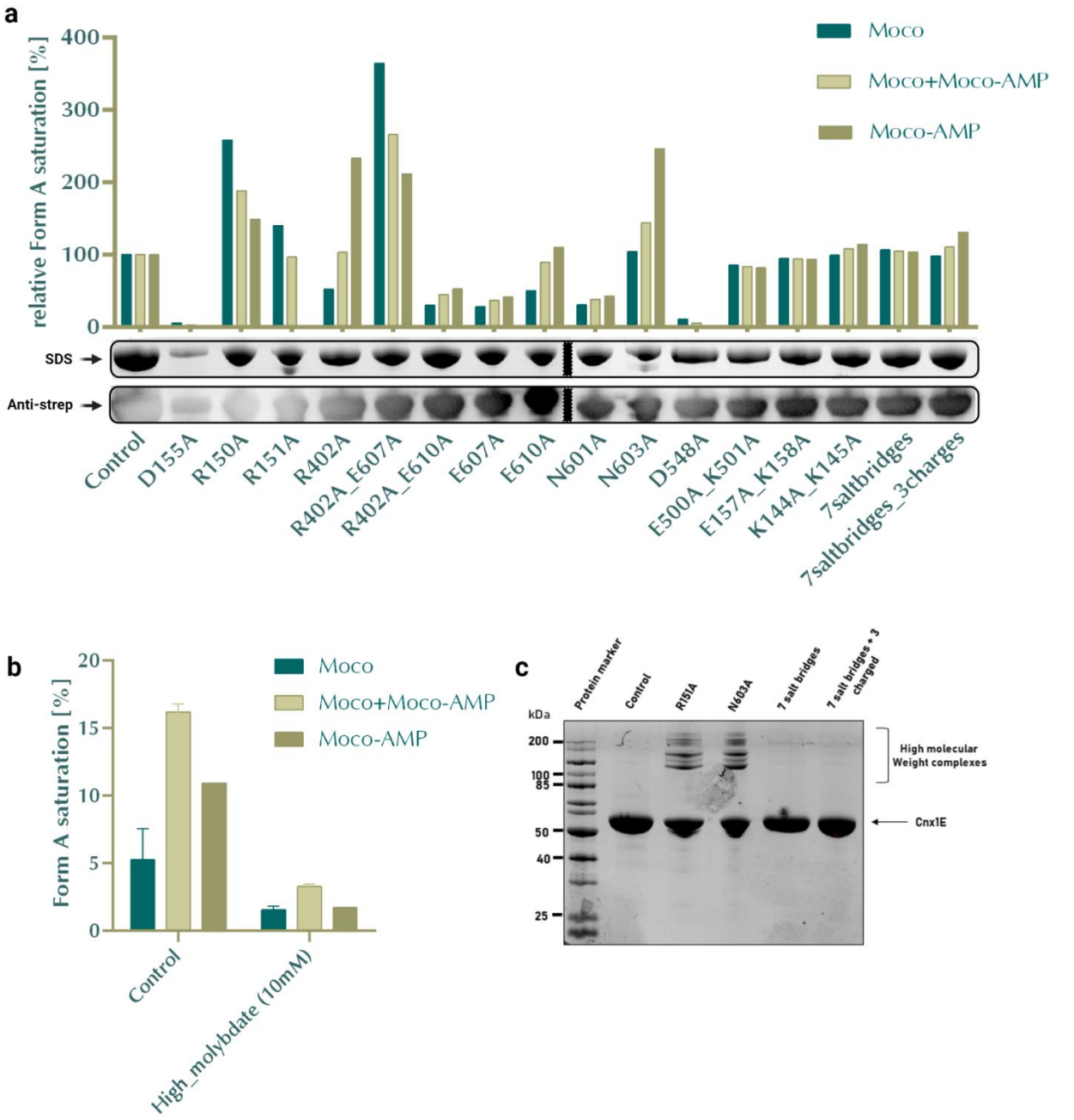



**Figure 52. Results of site-directed mutagenesis.** All results shown are of 3 technical replicates. (a) FormA-based detection results showing the relative saturations on the X-axis and the different mutants on the Y-axis, including S269D, D274S double mutant as control and D458A as a positive control. The control (Cnx1E variant S269D D274S) was set to 100%. All proteins where of comparable purity as documented for the control except for D155A. All proteins were checked with an SDS PAGE and western blotting for purity (Supplemental 69 and 70). For western blotting, an Anti-strep tag antibody was used (Ringel *et al.,* 2013). All bands showed at the size of ≈ 45 kDa that is equivalent to the size of Cnx1E. Additionally, both R151A and N603A showed high molecular weight bands. (M. Baldauf BA) (b) FormA-based detection results of S269D, D274S mutant used as a control with normal molybdate (1mM) and high molybdate concentration added (10 mM) during protein expression. (c)SDS PAGE of control, R151A, N603A and both 7 salt-bridges constructs. The SDS showed stabilized high molecular weight complexes without cross-linking in the case of R151A and N603A.

With the exception of D155A, R151A and N603A, SDS-PAGE and Anti-strep blotting results of the mutants showed identical behaviour to the control (Figure 52a). D155A showed abnormal minimal Cnx1E in the SDS gel with some degradation products. This was observed during the expression, which can be explained that D155A had improper protein folding. These results observed excludes D155A from the final analysis that would be incomparable to the other mutants and the controls used. For R151A and N603A, the SDS PAGE showed high molecular weight complexes that are stabilized on the denaturing gel without cross-linker addition (Figure 52c). Moreover, Anti-strep blotting showed these high molecular weight bands to be present. Disturbing single surface exposed residues showed to enhance complex formation (Tadokoro *et al.,* 2013).
To monitor the Moco-metabolites, a well-established quantification HPLC method was done (Krausze *et al.,* 2017; Krausze *et al.,* 2018) as described in section 2.2.3. All mutants were done as 3 technical replicates. Amongst the variants compared (orientation model 1 and 2) from the 16 tested variants, E607A, E610A, N601A, R402A_E610A turned out to be co-purified with a significantly reduced Moco-AMP content as documented by p values <0.001.

Analysis of the Cnx1G septuplet variant (7salt bridges) revealed that when co-expressed with Cnx1E variant S269D D274S the accumulating Moco-AMP amount is comparable to the control (Cnx1G wild type co-expressed with Cnx1E Moco-AMP accumulating variant S269D D274S). As co-expression of Cnx1E S269D D274S with the Cnx1G septuplet variant does not result in any measurable decrease in accumulating Moco-AMP, I conclude that this variant does not affect the MPT-AMP metabolite transfer (orientation 1). This finding is consistent with the fact that the interaction interface targeted here involves orientation 2. I further conclude that when expressed separately, E- and G-domain interact only via orientation 2. This finding is consistent with the fact that in prokaryotes, the structural proteomic approach only revealed a single interaction variant (similar to orientation 2). Assuming orientation 1 is essential for oligomerization and complex assembly of eukaryotic Mo-insertases, in bacteria, this orientation would not be necessary for the transfer of MPT-AMP from MogA to MoeA. This orientation may be for the interaction of Cnx1 with other interaction partners.

However, clearly it cannot be excluded, that the Cnx1G septuplet variant is still fully interaction competent with the E-domain. Subsequent studies could be carried out to clarify this.

The HPLC results showed a significantly reduced amount of relative saturation in the case of high molybdate (10 mM) compared to the standard control of 1mM. These results can confirm the hypothesis of two molybdate binding sites in Cnx1E, including the entry site and insertion site (Krausze *et al.,* 2018). In the case of normal molybdate amount, the Cnx1E binds molybdate initially at the entry site and afterwards at the insertion site. The crystal structures showed molybdate bound to Cnx1E in the two different binding sites, which showed that the presence of molybdate in the insertion site would not allow the binding of MPT-AMP to Cnx1E (Krausze *et al.,* 2018. In the case of high molybdate, the high influx of molybdate in Cnx1E surrounding would lead to a rapid turnover of molybdate from the entry site to the insertion site (low occupancy and high occupancy). This will result in hindering the binding of MPT-AMP to Cnx1E.

Next to Cnx1E / Cnx1G variants that resulted in an overall decrease of detectable Moco-AMP co-purified with Cnx1E S269D D274S, some variants were identified to result in an increase in Moco/MPT-AMP detectable. However, amongst these variants, high p values and correspondingly high standard deviations exclude – at least in most cases – a robust conclusion. Although the loss of a conserved salt bridge was shown to improve substrate binding (Mhaindarkar *et al.,* 2018), however, in the case of R402A_E607A and R151A, the high saturations shown are not reliable since both mutants had high standard deviation compared to the control.

### 3.3.4 Summary

In summary, the results presented of the interaction models (orientation 1 and 2) were used to generate a set of residues potentially essential for E- and G-domain interaction, which was targeted by structure-guided mutagenesis. Unexpectedly, none of the variants targeting interaction model 1 (orientation 1) resulted in a decrease of detectable, Cnx1E bound Moco-AMP. Hence it was evident that this interaction model plays no role when both domains are expressed separately. On the contrary, when targeting interaction model 2, it became evident that some residues (i.e. E-domain: R402, G-domain E607 and E610) play an essential role in the functional interaction of both domains, as documented by the finding that Cnx1E was co-purified with significantly reduced Moco-AMP saturations.

## 3.4. Gephyrin

### 3.4.1.Overview

Whilst the prokaryotic Mo-insertases assemble from separately expressed E- and G-domains, eukaryotic E- and G-domains are fused, hence forming a single multi-domain protein. As a major difference between plant and fungal Mo-insertases, mammalian Mo-insertases possess an enlarged linker region, which is considered as a third domain (Schwarz *et al.,* 2000). Mammalian Mo-insertases have a central role in the cellular Mo-metabolism essential for the clustering of GABA receptors in the post-synapsis. This function however, is independent of G-domain but is associated with the E-domain or C-domain (linker) of gephyrin (Choii & Ko, 2015). In the cell, different gephyrin variants exist from which recent work identified a variant being involved in the Mo-insertion reaction. In the following, I recombinantly expressed this variant and used the pure protein preparations obtained for an itial cryo-EM approach.

### 3.4.2. Expression and purification of gephyrin

The full-length *Homo sapien* gephyrin (splice variant P2) was recombinantly expressed and purified from *E. coli* according to the established protocols for recombinant Cnx1 full-length production shown in section 3.2.2 (Figure 27). SDS PAGE was done to monitor the purity of the purification (Figure 53). Next to human gephyrin, I also recombinantly expressed and purified R. norvegicus full-length gephyrin for comparison. Splice varient P1 and P2 of R. norvegicus were purified as a comparison. In general, both splice variants, P1 and P2, are involved in the Moco biosynthesis, and other splice variants do not have Moco activity and dictate other neuronal functions (Tyagarajan & Fritschy, 2014). For this reason, P1 or P2 is the most commonly used gephyrin splice variants when studying Moco and Mo-insertases (Babu & Schindelin, 2018).

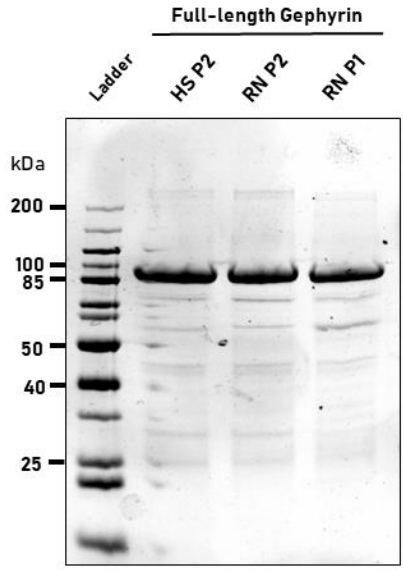



**Figure 53 SDS PAGE of recombinant full-length gephyrin from Homo sapiens and Rattus norvegicus.** Full-length proteins were purified from *Homo sapiens* isoform 2 (HS P2), *Rattus norvegicus* isoform 2 (RN P2) and *Rattus norvegicus* isoform 1 (RN P1). The bands showed at the expected band size of approximately 90 kDa for all three proteins. (Gephyrin results are prepared with the help of M. Baldauf and J. Plate, BA Milena Baldauf)

The general expression conditions were initially adopted from previously published work (Babu *et al.,* 2019). Nonetheless, expression conditions were checked in order to optimize the purification of the full-length proteins with minimal degradation products. To evaluate the expression conditions, various expression parameters were conducted, including different IPTG concentrations and temperatures. Afterwards, cell lysate was loaded on SDS-PAGE (Figure 54). Initially, gephyrin was expressed with 0.5 mM IPTG at 16 C◦ overnight. IPTG concentration was optimized by using 3 different IPTG concentrations including 0.1 mM, 0.5 mM and 1 mM. Additionally, different expression temperatures include maintaining a 37 C◦ for 3 hrs after induction to express the protein. As a result, the SDS-PAGE showed that after 3 hrs of induction, at 16 ◦C, no protein is expressed, however at 37 ◦C, full-length proteins are observed as well as high molecular weight complexes with minimal degradation, regardless of IPTG concentration (higher IPTG concentration showed complexes while 0.1 mM did not show complexes). On the next day at 16 ◦C (the conditions used in this work), minimal degradation products are observed with high molecular weight bands above 200 kDa. The higher the IPTG concentration, the more complex is observed.

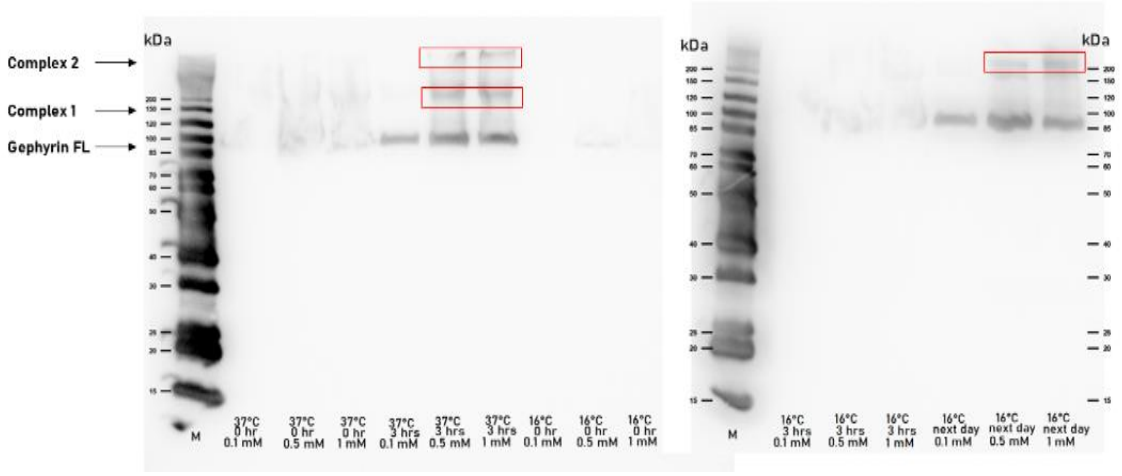



**Figure 54 Anti-strep western blot of the various expression parameters tested for recombinant gephyrin expression.** The western blot was prepared by loading lysate from different conditions. Three different IPTG concentrations were added to each condition, including 0.1 mM, 0.5 mM, and 1 mM. Samples were taken at 0 hr, 3 hrs and the next day after induction. The temperature was kept at 37 ◦C or decreased to 16 ◦C. The results show minimal protein expression in general at 0 hrs induction. After 3 hrs of induction, at 16 ◦C, no protein is expressed, however at 37 ◦C, full-length protein is observed as well as high molecular weight complexes with minimal degradation regardless of IPTG concentration. At 16 ◦C after overnight cultivation, minimal degradation products were observed with high molecular weight bands above 200 kDa.

## 3.4.3. Crosslinking gephyrin

In order to stabilize the E-G domain interactions within gephyrin, the samples were routinely cross-linked with DSS prior to negative staining EM. For cross linking, I used DSS´due to the similarity of chemical structure and reactivity to DSBU, DSS was used to monitor the successfulness of the crosslinking of the human gephyrin. Upon crosslinking, SDS PAGE analysis revealed high molecular weight complexes of ≈ 300 kDa (Figure 55).

Notably, crosslinking of MoeA-MogA and Cnx1 likewise resulted in formation of complexes with comparable molecular weights (see figure 20 and 32). Additionally, anti-Nit-9E (E-domain of *N. crassa* Mo-insertase) blots of non-crosslinked *N. crassa* extract showed the same high molecular weight complexes (Supplemental figure 77 and 78).

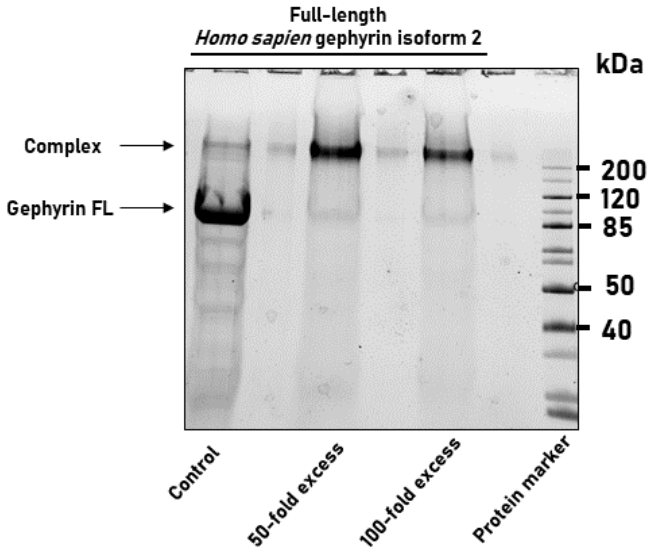



**Figure 55 . SDS PAGE of cross-linked recombinant Homo sapiens gephyrin isoform 2.** DSS was used to cross-link gephyrin. Therefore, a 50- and 100-fold excess of the cross-linker was applied. Full-length gephyrin was detected as a 90 kDa sized protein. As documented in the lane I (control), a higher molecular weight gephyrin complex was readily present in the non-cross linked protein preparation. Upon cross-linking, the full-length gephyrin was hardly detectable, while the higher molecular weight complex became clearly detectable.

## 3.4.4. Negative staining of gephyrin

As described in 2.2.8, gephyrin was negative stained and particles were manually selected with a mask dimeter of 250 Å circular range. Auto-picking with a 200 Å range and a 2D classification was done with limiting of resolution to 30 Å for 50 classes. Surprisingly, compared to the MoeA-MogA complex and Cnx1, gephyrin showed more homogeneity. This made 3D modelling easier as a better data set could be obtained. The data was then processed further, and the first models were generated based on the negative staining results.

## 3.4.5. Modelling

For 3D modelling of gephyrin, 3 to 5 sub-classes we selected, and an initial 3D model was done using a mask diameter of 220 Å. All models were visualized using Chimera (RBVI). The models generated are shown in figure 56 and 57.

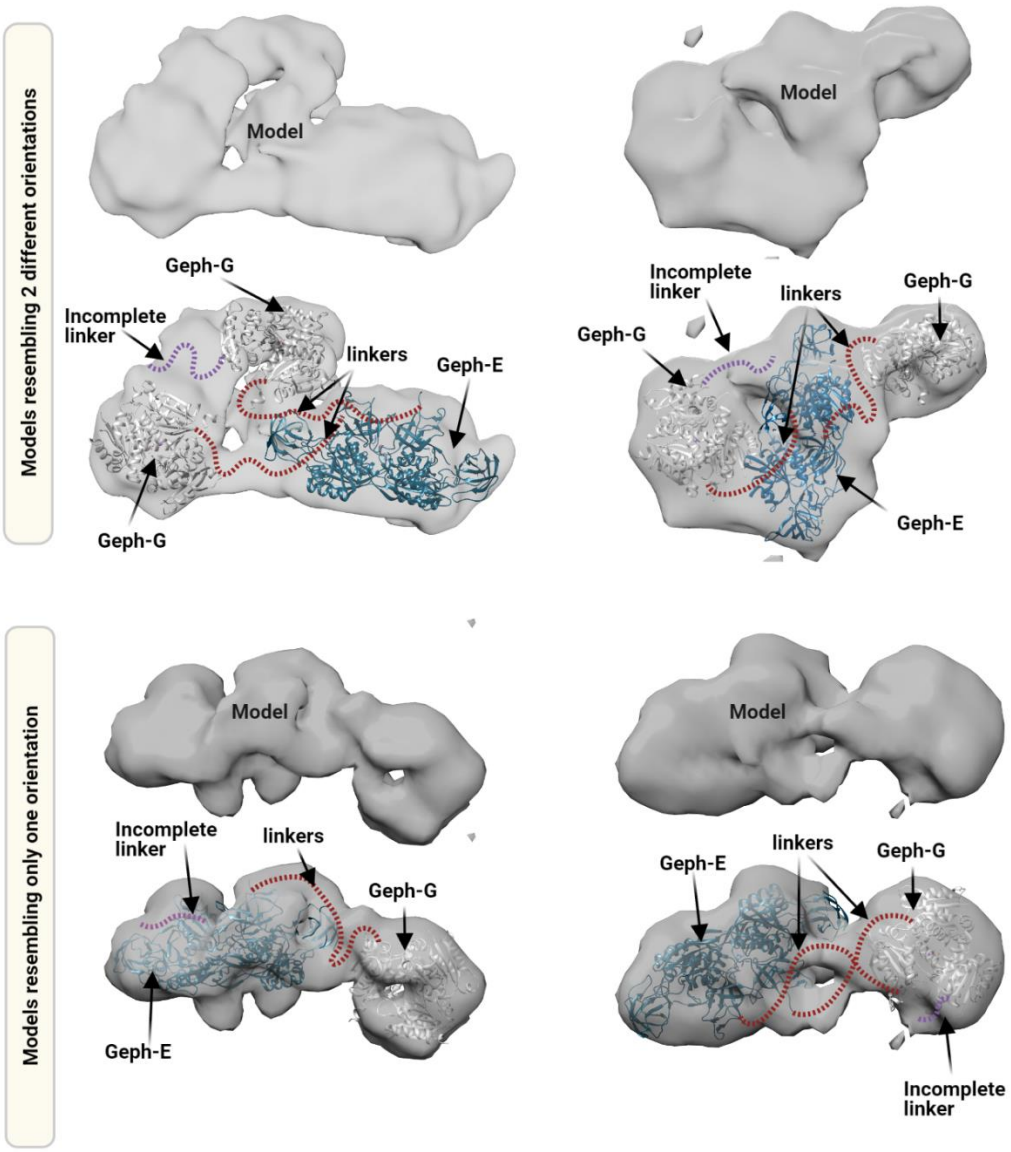



**Figure 56 Modelling gephyrin from the classification of the negative staining.** The four models shown are based on different 2D classes selected for 3D modelling. Each model was generated based on various sub-classes with similar shapes. The structures used for modelling were PDB 1JLJ for Geph-G (white) and PDB 4TK3 for Geph-E (dark blue). The linker is schematically shown in red, while incomplete linkers are shown in purple. The linkers were drawn but not calculated based on the estimated length of the sequence of the linker in Å. All structures were fitted in the envelope map derived from cryo-EM using Chimera. The top two models resemble the two orientation model from the Cnx1 complex, and the lower two models resemble one orientation only from the Cnx1 model generated.

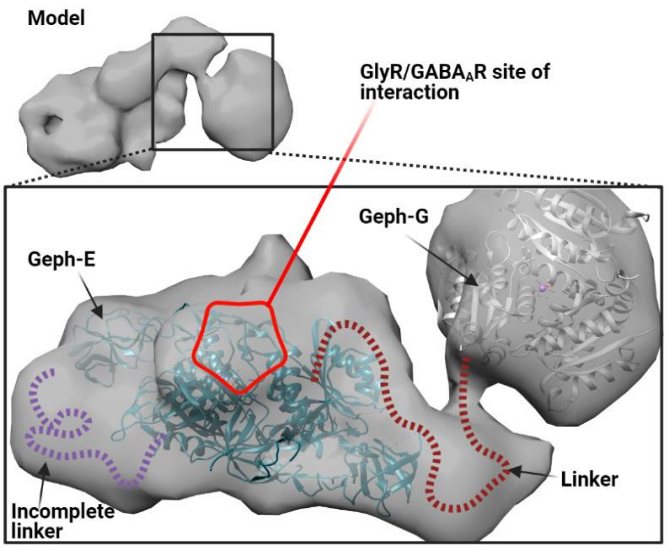



**Figure 57 . 3D model of gephyrin highlighting GlyR/GABAA site of interaction.** The model presented is based on different 2D classes selected for 3D modelling. Each model was generated based on various sub-classes with similar shapes. The binding site of the glycine receptor (GlyR) and $GABA_A$ receptor are highlighted. The structures used for modelling were PDB: 1JLJ for Geph-G (white) and PDB: 4TK3 for Geph-E (Dark blue). The linker was schematically represented in red, while incomplete linkers were represented in purple in the vacant space of the envelope. All structures were fitted in an envelope map using Chimera (RBVI).

Using Robetta as described in 2.2.7., the gephyrin linker was modelled *de novo* (Figure 58). The result presented five models of which the top two models were selected for further analysis. The two models were selected based on the modelling scores and the manual assessment done by checking the similarity of the structured regions (E- and G- domain) with the overall model shape. The final models resembled the two orientations resulting from the Cnx1 modelling. However, for gephyrin, the long linker gives the G-domain (Geph-G) further flexibility and movement when interacting with the E-domain (Geph-E) as the model shows in figure 58. This movement allows the binding of E-domain to G-domain in a relaxed manner compared to the shorter linker in plants.

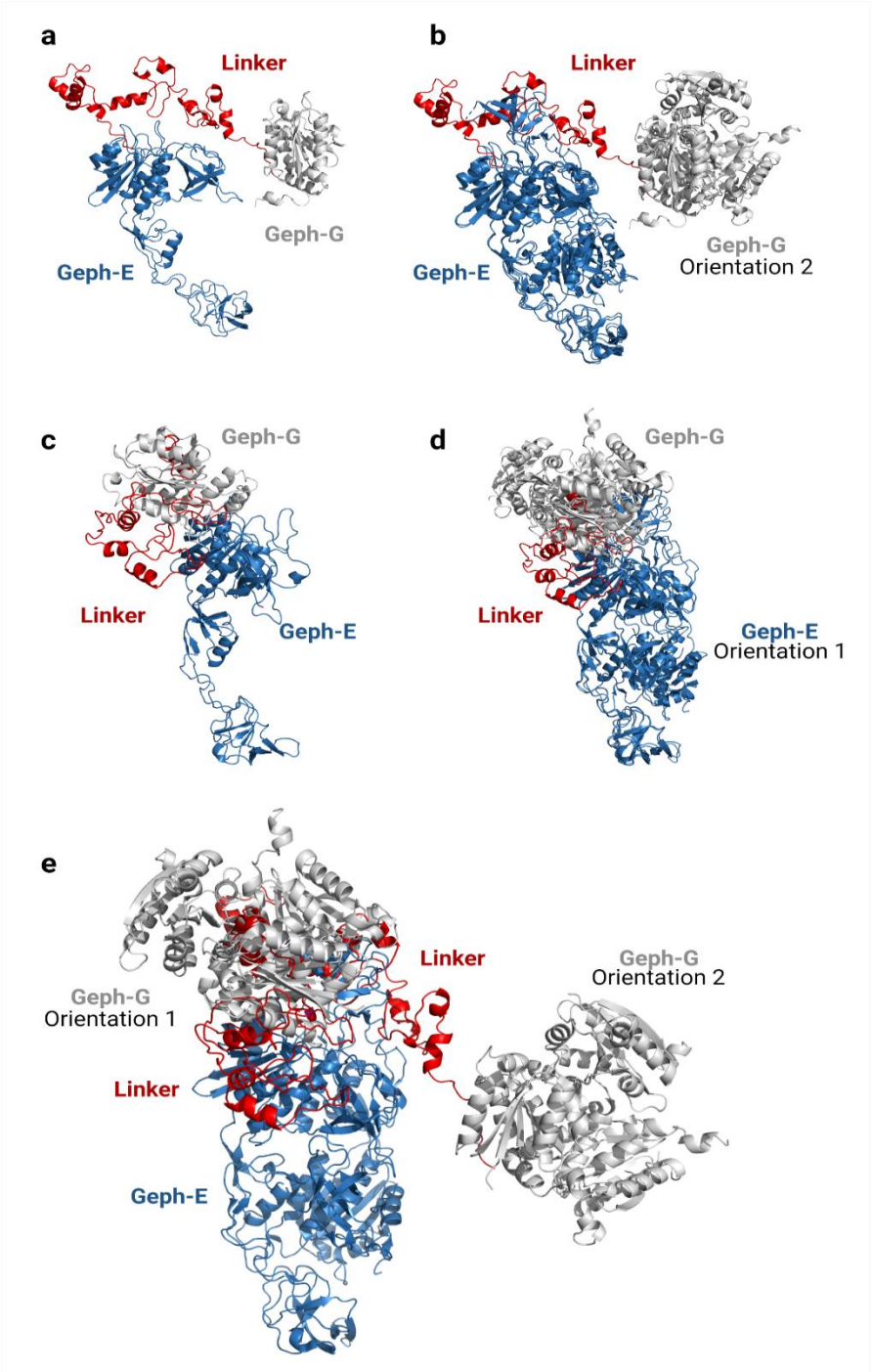



**Figure 58 Modelling the linker of full-length gephyrin.** Modelling was done using Robetta (Simons *et al.,* 1999). The modelling included Geph-E (E- domain, blue), Geph-G (G-domain, grey) and the linker (C-domain, red). Superimposition of the *ab initio* modelled linker was done using PDB: 4TK3 (Geph-E) and PDB: 1JLJ (Geph-G). (a) The first model of gephyrin full-length. (b) Superimposition of Geph-E dimer and Geph-G trimer over the first Robetta model (resembling orientation 2). (c) The second model of gephyrin full-length. (d) Superimposition of Geph-E dimer and Geph-G trimer over the second Robetta model (resembling orientation 1). (e) Overly of both models generated.

The capability of gephyrin to synthesize Moco more efficiently in full-length than the separate domains (Belaidi & Schwarz *et al.* 2013) hints at the evolutionary importance of this fusion to give novel functions. Interestingly, a loop that is surface-exposed in the E-domain regulates postsynaptic clustering in an absolute manner (Lardi-Studler *et al.,* 2008; Meyer *et al.,* 1995). The binding site for GlyR β-loop is on the E-domain as well. The model shown gives the possibility for the interaction with other proteins and GlyR due to the possible access (Figure 57). The models shown in figure 56 and 57 are more similar to orientation 2 in Cnx1 rather than orientation 1 (in agreement with the modelling of the linker). Since all interactions know with gephyrin are mainly with the E-domain and not the G-domain, the G-domain is for the sole function of Mo-insertion and possibly structural reasons. With a long flexible linker, a dynamic-flexible complex is expected (Marsh & Teichmann, 2015). Notably, the addition of the cross-linker constricts the possible complex flexibility leading to the presence of further confirmations of the complex (Bullock *et al.,* 2018). This explains the problematic particle classification with negative staining EM that is inevitable in the gephyrin samples. The linker showed in figure 56 and 57 was illustrated in the approximate distance with the *ab initio* gephyrin modelled linker shown in figure 58. Due to the low resolution of the models, it is hard to explain the vacant space around both domains and the linker. This could be avoided when freezing gephyrin samples and applying Cryo-EM, which would generate a superior dataset for model building.

### 3.4.6 Summary

Next to the prokaryotic Mo-insertase MoeA/MogA and the plant Mo-insertase Cnx1 also the human Mo-insertase gephyrin was characterized within this work. For doing so, no XL-MS based approach was employed, however, the highly pure full-length protein was successfully used for negative staining EM, which was used for subsequent 3D reconstruction modelling.

As a major finding, the two G-domain binding sites were also identified for the GephE-domain, hence rendering also this interaction model to comprise both functional (metabolite transfer) and structural aspects.

## 3.5. Interaction of Mo-insertase with Mo-dependent enzyme

Hitherto, I focused on the assembly of E- and G-domain within the Mo-insertase complex. However, upon synthesis, Moco requires the protein protected transfer either to a carrier protein or to the user enzyme directly (Kruse *et al.,* 2010; Mendel 2013 Krausze *et al.,* 2020; Hercher *et al.,* 2020). Previous work readily showed the interaction of recombinant Cnx1E with *A. thaliana* Moco binding proteins (Kruse *et al.,* 2010) and *in vivo* work complemented these data sets (Kaufholdt *et al.,* 2017).

In my work, I used the *N. crassa* system to test for the directed interaction of its Mo-insertase (Nit-9) with the major user of Moco in *N. crassa* which is the nitrate reductase (Wajmann *et al.,* 2020). To do so, I carried a fully defined *in vitro* approach, and I used recombinant Nit-9 and the Moco-dimer domain (Ringel *et al.,* 2013) of the *N. crassa* NR. In the following, I describe the results obtained for the fully defined *in vitro* approach.

To investigate direct Mo-insertase interaction with user enzymes, full-length Nit-9 (*N. crassa* Mo-insertase) was recombinantly expressed and purified as described in 2.2.2.1 and 2.2.2.2 with the protocol generated in this work. The recombinant full-length proteins were then tested with Nit-9E anti-bodies (Ringel *et al.,* 2015) (Supplemental figure 78). The recombinant Nit-9 full-length protein was *in vitro* cross-linked with the Moco-binding domain of *N. crassa*'s NR (expressed and purified as full-length protein for the first time in this work). Therefore, equimolar amounts of both proteins were co-incubated for 60 min at room temperature prior to adding the 100-fold excess of the DSS-cross linker (Crosslinking described in 2.2.4).

As a control, cross-linking was done for the individual proteins. Next to this, *A. thaliana* Cnx1 full length was used for cross-linking following the experimental setup described for Nit-9 (See 2.2.2.2). Doing so did not reveal Cnx1-NR complex-formation. However, full-length Nit-9 and NR were identified to form two complexes with estimated molecular weights of ≈250 kDa and ≈ 400 kDa, respectively.

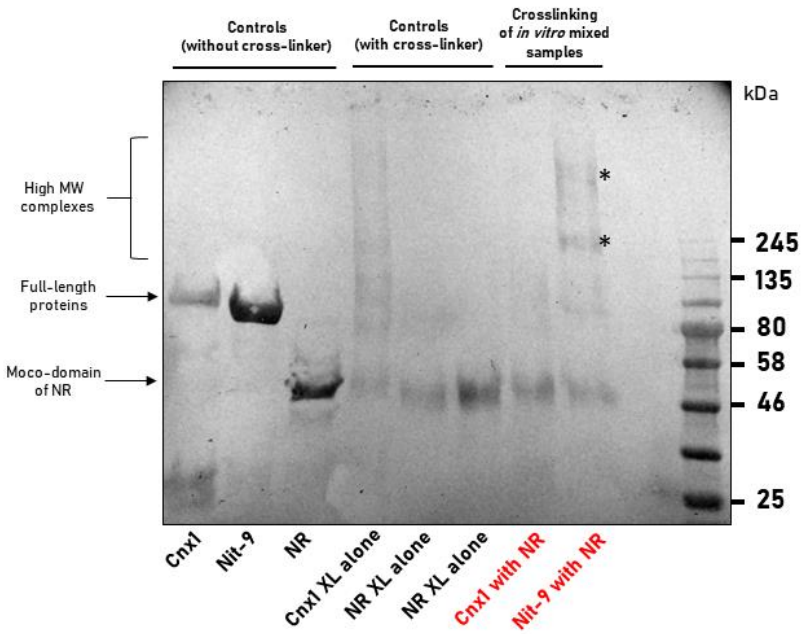



**Figure 59 Analysis of *N. crassa* NR interactions with Cnx1 and Nit-9.** The blot shown was stained with Ponceau S. DSS was used for all crosslinking with a 100-fold excess of the cross-linker to the protein. Full-length proteins were detected with a ≈ 90 kDa size. Controls were loaded in the first (Cnx1), second (Nit-9) and third lane (Moco binding domain of NR). Cnx1 crosslinking with NR did not show any high molecular weight complexes. However, Nit-9-NR crosslinking resulted in high molecular weight complexes of ≥250 kDa (marked with asterisks).

This interaction is selective, at least for *N. crassa*, and cannot be complemented with the Mo-insertase of *A. thaliana*. Additionally, the binding site of the Mo-insertase would be, as expected, in the binding site of Moco in the NR. There is no crystal structure yet available for Nit-9 individual domains. NR would bind to the E-domain of the Mo-insertase as it contains the substrate for NR. The crosslinking result shows selectivity in terms of binding of NR to the Mo-insertase of *N. crassa*.

Speculatively, in an initial approach to give a possible explanation for this finding, an amino acid sequence alignment was carried out using various plant E- domains and Nit-9. Hypothetically, Nit-9E was found to contain a sequence stretch comprising residues 117 to 128, which is lacking in any of the compared plant E-domains. As structural information for Nit-9E is missing, I carried out a PHYRE2-based modelling approach, which revealed a sequence stretch to be located in subdomains I+II (see Fig.60). However, another possible binding site may be formed by a non-conserved *N. crassa* NR sequence stretch, which is not present in other plant species (Figure 60). The sequence stretch is located in the Moco-

binding domain of *N. crassa* NR. However, based on the available results, none of the two models can be excluded.

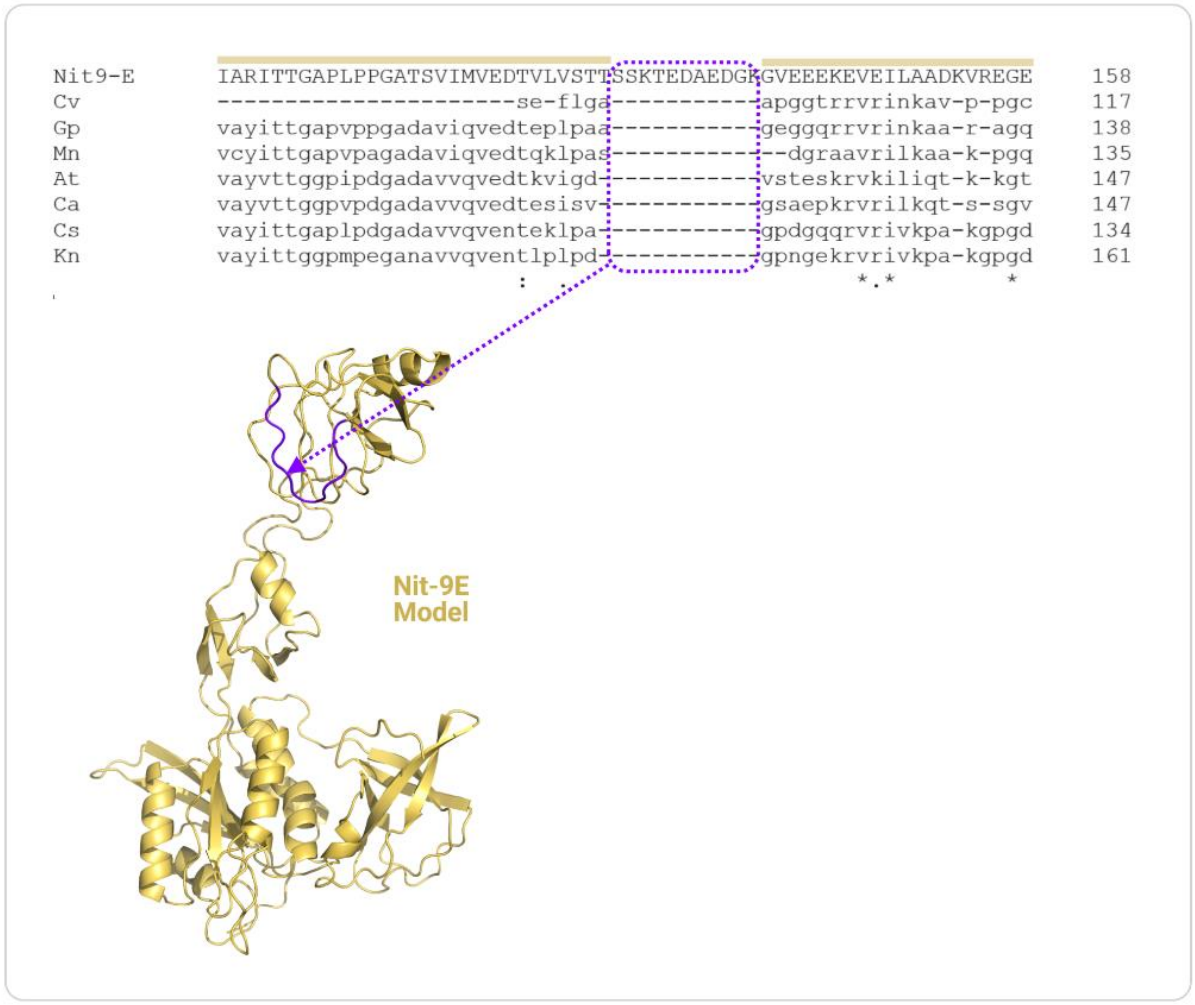



**Figure 60 . *N. crassa* Nit-9E specific sequence stretch.** (a) Partial amino acid sequence alignment of plant Mo-insertases from Cv, (*Chlorella variabilis*), Gp (*Gonium pectoral*), Mn (*Monoraphidium neglectum*), At (*Arabidopsis thaliana*), Ca (*Capsicum* annum), Cs (*Coccomyxa* subellipsoidea) and Kn (*Klebsormidium nitens*) with *Neurospora crassa* (Nit-9E) and (b) Phyre2 based prediction of the *Neurospora crassa* Nit9-E structure.

## 3.6. Overall conclusion

In this work, I elucidated the principles behind Mo-insertase complex assembly of different eukaryotic and one prokaryotic species. This was possible by combining different techniques, i.e. structure-based proteomics, cryo-EM, native mass spectrometry and recombinant biochemistry. As a major finding, a common mode of how E- and G-domains are assembled in the Mo-insertase complexes of prokaryotes and eukaryotes was suggested to exist. In this complex, the active sites of both G- and E-domain face each other. Structure-based evidence suggested that the metabolite transfer between both domains involves the reversible movement of one flexible subdomain found in all prokaryotic and eukaryotic E-domains. Therefore, I conclude that the primary mechanism of metabolite transfer between G- and E-domains of prokaryotic and eukaryotic Mo-insertases is conserved. This finding is consistent with earlier work from the Mendel group (Stallmeyer et al., 1999), which identified the E- and G-domain associated Mo-insertase activity to be complemented between different species (Figure 61).

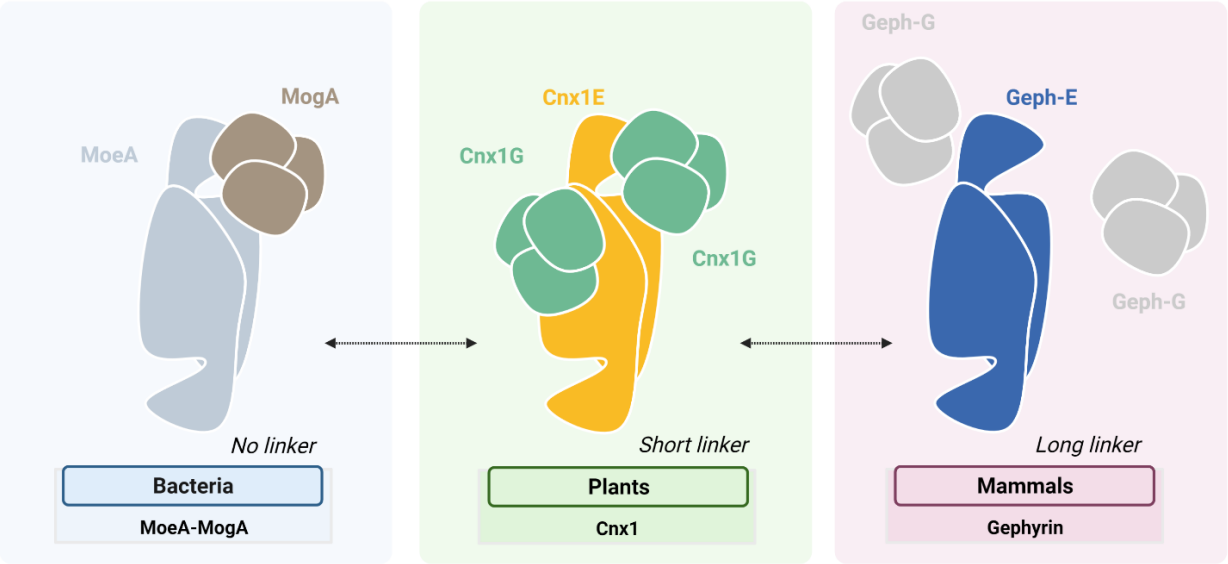



**Figure 61 . Schematic representation of the Mo-insertase complex in different species.** Mo-insertase of bacteria composed of MoeA dimer (grey) bound to one MogA trimer (brown) in orientation 2 manner. For plants, Mo-insertases composed of two orientation of the G-domain presented in this work as two trimers (teal) bound to one Cnx1E dimer (bright orange). In mammals, the gephyrin E- domain (blue) interacts with G- domain (grey). MoeA-MogA does not contain a linker between both E-G domains. Cnx1 has a short linker, while gephyrin contains a long linker. All Mo-insertases were shown to be complemented, as pointed out in dotted arrows (Stallmeyer *et al.,* 1999).

Notably, next to the universally present E-G domain assembly within the Mo-insertase complex, the E-domain possesses a second G-domain binding site in eukaryotes. This in turn, allows the formation of a bigger sized complex (comprising 3 E-domains and 2 G-domains). However, using this complex as a building unit, in principle also complexes consisting of 6 E-domains and 4 G domains could exist. Accepting that *in cellula* Cnx1G trimers exist (as a result of Cnx1 full length degradation), would even allow the formation of complexes with other stoichiometries of E- and G-domains.

I assume that the formation of Mo-insertase complexes containing non-Cnx1E linked trimeric G-domains is beneficial for the cellular Moco-synthesis activity, as additional MPT-synthesis capacity would be incorporated into the complex. As a direct consequence, this would enhance the rate of Moco-synthesis. Assuming this to be true, I conclude that the formation of the eukaryotic Mo-insertase complex is a dynamic and potentially also a demand-driven process.

Potentially, the linker length impacts complex assembly and dynamics, as naturally, a prolonged linker region will allow for variations of the E-G domain assembly. Thus, altered assembly units may subsequently result in other E-/G-domain stoichiometries and orientations within the Mo-insertase complex.

The results of my work provide the first structural insights into the overall architecture of the Mo-insertase complex, which may serve as a groundwork upon which future structural and biochemical studies may continue to build.

## 3.7. Outlook and further research

The results shown here have opened the door to further biochemical and structural studies on Mo-insertase complexes. Further exploration into the structure of Cnx1 and gephyrin can explain the exact structural mechanism of the linker. Importantly, biochemical approaches can identify how exactly does the linker impact the metabolic transfer. Further Cryo-EM data, with the support of biochemical data, from the Mo-insertases can give us more detailed information about the complex. With the support of the data here, it is possible to visualize the dynamic motion of how MPT-AMP is precisely transferred from Cnx1G to Cnx1E by further structural and biochemical experiments.

Moreover, since it is easier to study larger complexes with Cryo-EM, Cnx1 and MPT-synthase complex can be envisioned. Direct interaction of Nit-9 or Cnx1 with the Nitrate reductase can be structurally investigated with the information that is presented in this work. From the neuroscientific perspective, the overall shape of the complex can also be used as a platform for structural input with the other interactors, such as the actin, glycine and GABAA receptors.

# References

A. Havemeyer, J. Lang, B. Clement, The fourth mammalian molybdenum enzyme mARC: current state of research, Drug Metab. Rev. 43 (2011) 524–539

A. Llamas, T. Otte, G. Multhaup, R.R. Mendel, G. Schwarz, The mechanism of nucleotide-assisted molybdenum insertion into molybdopterin. A novel route to- ward metal cofactor assembly, J. Biol. Chem. 281 (2006) 18343–18350.

Allison, D. W., Chervin, A. S., Gelfand, V. I., & Craig, A. M. (2000). Postsynaptic scaffolds of excitatory and inhibitory synapses in hippocampal neurons: Maintenance of core components independent of actin filaments and microtubules. Journal of Neuroscience, 20(12), 4545–4554. https://doi.org/10.1523/jneurosci.20-12-04545.2000

Arlt, C., Ihling, C. H., & Sinz, A. (2015). Structure of full-length p53 tumor suppressor probed by chemical cross-linking and mass spectrometry. Proteomics, 15(16), 2746–2755. https://doi.org/10.1002/pmic.201400549

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T., & Ben-Tal, N. (2016). ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Research, 44(W1), W344–W350. https://doi.org/10.1093/nar/gkw408

Babu, M., Bundalovic-Torma, C., Calmettes, C., Phanse, S., Zhang, Q., Jiang, Y., Minic, Z., Kim, S., Mehla, J., Gagarinova, A., Rodionova, I., Kumar, A., Guo, H., Kagan, O., Pogoutse, O., Aoki, H., Deineko, V., Caufield, J. H., Holtzapple, E., … Emili, A. (2018). Global landscape of cell envelope protein complexes in Escherichia coli. Nature Biotechnology, 36(1), 103–112. https://doi.org/10.1038/nbt.4024

Belaidi, A. A., & Schwarz, G. (2013). Metal insertion into the molybdenum cofactor: Product-substrate channelling demonstrates the functional origin of domain fusion in gephyrin. Biochemical Journal, 450(1), 149–157. https://doi.org/10.1042/BJ20121078

Blanchet, C. E., & Svergun, D. I. (2013). Small-angle X-ray scattering on biological macromolecules and nanocomposites in solution. Annual Review of Physical Chemistry, 64, 37–54. https://doi.org/10.1146/annurev-physchem-040412-110132

Bosshard, H. R., Marti, D. N., & Jelesarov, I. (2004). Protein stabilization by salt bridges: Concepts, experimental approaches and clarification of some misunderstandings. Journal of Molecular Recognition, 17(1), 1–16. https://doi.org/10.1002/jmr.657

Bricogne G, B. E, Brandl M, Flensburg C, Keller P, Paciorek W, Roversi P, Sharff A, Smart OS, Vonrhein C, Womack TO (2017) BUSTER version 2.10.3 Cambridge, United Kingdom: Global Phasing Ltd.

Brusotti, G., Calleri, E., Colombo, R., Massolini, G., Rinaldi, F., & Temporini, C. (2018). Advances on Size Exclusion Chromatography and Applications on the Analysis of Protein

Biopharmaceuticals and Protein Aggregates: A Mini Review. Chromatographia, 81(1), 3–23. https://doi.org/10.1007/s10337-017-3380-5

Bui, K. H., Von Appen, A., Diguilio, A. L., Ori, A., Sparks, L., Mackmull, M. T., Bock, T., Hagen, W., Andrӣs-Pons, A., Glavy, J. S., & Beck, M. (2013). Integrated structural analysis of the human nuclear pore complex scaffold. Cell, 155(6), 1233–1243. https://doi.org/10.1016/j.cell.2013.10.055

Bullock, J. M. A., Sen, N., Thalassinos, K., & Topf, M. (2018). Modeling Protein Complexes Using Restraints from Crosslinking Mass Spectrometry. Structure, 26(7), 1015-1024.e2. https://doi.org/10.1016/j.str.2018.04.016

Chivian, D., Kim, D. E., Malmstrum, L., Bradley, P., Robertson, T., Murphy, P., Strauss, C. E. M., Bonneau, R., Rohl, C. A., & Baker, D. (2003). Automated Prediction of CASP-5 Structures Using the Robetta Server. Proteins: Structure, Function and Genetics, 53(SUPPL. 6), 524–533. https://doi.org/10.1002/prot.10529

Choii, G., & Ko, J. (2015). Gephyrin: a central GABAergic synapse organizer. Experimental & Molecular Medicine, 47(November 2014), e158. https://doi.org/10.1038/emm.2015.5

Cohen HJ, Fridovich I, Rajagopalan KV. Hepatic sulfite oxidase. A functional role for molybdenum. J Biol Chem. 1971 Jan 25;246(2):374-82. PMID: 5100417.

Dau, T., Gupta, K., Berger, I., & Rappsilber, J. (2019). Sequential Digestion with Trypsin and Elastase in Cross-Linking Mass Spectrometry. Analytical Chemistry, 91(7), 4472–4478. https://doi.org/10.1021/acs.analchem.8b05222

de Vries, S. J., & Bonvin, A. M. J. J. (2011). Cport: A consensus interface predictor and its performance in prediction-driven docking with HADDOCK. PLoS ONE, 6(3). https://doi.org/10.1371/journal.pone.0017695

Desta, I. T., Porter, K. A., Xia, B., Kozakov, D., & Vajda, S. (2020). Performance and Its Limits in Rigid Body Protein-Protein Docking. Structure, 28(9), 1071-1081.e3. https://doi.org/10.1016/j.str.2020.06.006

E. Bayram, Y. Topcu, P. Karakaya, U. Yis, H. Cakmakci, K. Ichida, S.H. Kurul, Molybdenum cofactor deficiency: review of 12 cases (MoCD and review), Eur. J. Paediatr. Neurol. 17 (2013) 1–6.

E. Krissinel and K. Henrick (2007). 'Inference of macromolecular assemblies from crystalline state.'. J. Mol. Biol. 372, 774--797.

Ellis, R. J. (2001). Macromolecular crowding: An important but neglected aspect of the intracellular environment. Current Opinion in Structural Biology, 11(1), 114–119. https://doi.org/10.1016/S0959-440X(00)00172-X

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica Section D: Biological Crystallography, 66(4), 486–501. https://doi.org/10.1107/S0907444910007493

Food and Drug Administration. (2016). Nutrition Facts Label. https://s3.amazonaws.com/public-inspection.federalregister.gov/2016-11867.pdf

Fritschy JM, Harvey RJ, Schwarz G. Gephyrin: where do we stand, where do we go? Trends Neurosci. 2008 May;31(5):257-64.

G. Schwarz, Molybdenum cofactor and human disease, Curr. Opin. Chem. Biol. 31 (2016) 179–187.

Gutze, M., Iacobucci, C., Ihling, C. H., & Sinz, A. (2019). A Simple Cross-Linking/Mass Spectrometry Workflow for Studying System-wide Protein Interactions. Analytical Chemistry, 91(15), 10236–10244. https://doi.org/10.1021/acs.analchem.9b02372

Groeneweg, F. L., Trattnig, C., Kuhse, J., Nawrotzki, R. A., & Kirsch, J. (2018). Gephyrin: a key regulatory protein of inhibitory synapses and beyond. Histochemistry and Cell Biology, 150(5), 489–508. https://doi.org/10.1007/s00418-018-1725-2

Hashimoto, K., & Panchenko, A. R. (2010). Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. Proceedings of the National Academy of Sciences of the United States of America, 107(47), 20352–20357. https://doi.org/10.1073/pnas.1012999107

Hercher, T. W., Krausze, J., Hoffmeister, S., Zwerschke, D., Lindel, T., Blankenfeldt, W., Mendel, R. R., & Kruse, T. (2020). Insights into the Cnx1E catalyzed MPT-AMP hydrolysis. Bioscience Reports, 40(1), 1–12. https://doi.org/10.1042/BSR20191806

Hille, R., Hall, J., & Basu, P. (1996). The mononuclear molybdenum enzymes. Chemical Reviews, 114(7), 3963–4038. https://doi.org/10.1021/cr400443z

Hou, Q., Bourgeas, R., Pucci, F., & Rooman, M. (2018). Computational analysis of the amino acid interactions that promote or decrease protein solubility. Scientific Reports, 8(1), 1–13. https://doi.org/10.1038/s41598-018-32988-w

Hrabe, T., Li, Z., Sedova, M., Rotkiewicz, P., Jaroszewski, L., & Godzik, A. (2016). PDBFlex: Exploring flexibility in protein structures. Nucleic Acids Research, 44(D1), D423–D428. https://doi.org/10.1093/nar/gkv1316

Huang, W., Masureel, M., Qu, Q., Janetzko, J., Inoue, A., Kato, H. E., Robertson, M. J., Nguyen, K. C., Glenn, J. S., Skiniotis, G., & Kobilka, B. K. (2020). Structure of the neurotensin receptor 1 in complex with β-arrestin 1. Nature, 579(7798), 303–308. https://doi.org/10.1038/s41586-020-1953-1

Iacobucci, C., Gutze, M., Ihling, C. H., Piotrowski, C., Arlt, C., Schäfer, M., Hage, C., Schmidt, R., & Sinz, A. (2018). A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. Nature Protocols, 13(12), 2864–2889. https://doi.org/10.1038/s41596-018-0068-8

Ihling, C. H., Springorum, P., Iacobucci, C., Hage, C., Gutze, M., Schäfer, M., & Sinz, A. (2020). The isotope-labeled, MS-cleavable cross-linker disuccinimidyl dibutyric urea for improved cross-linking/mass spectrometry studies. Journal of the American Society for Mass Spectrometry, 31(2), 183–189. https://doi.org/10.1021/jasms.9b00008

Iobbi-Nivol, C., & Leimkьhler, S. (2013). Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in Escherichia coli. Biochimica et Biophysica Acta - Bioenergetics, 1827(8–9), 1086–1101. https://doi.org/10.1016/j.bbabio.2012.11.007

J. Reiss, R. Hahnewald, Molybdenum cofactor deficiency: mutations in GPHN, MOCS1, and MOCS2, Hum. Mutat. 32 (2011) 10–18.

K. Ichida, T. Matsumura, R. Sakuma, T. Hosoya, T. Nishino, Mutation of human molybdenum cofactor sulfurase gene is responsible for classical xanthinuria type II, Biochem. Biophys. Res. Commun. 282 (2001) 1194–1200.

K. Yokoyama, E.A. Lilla, C-C bond forming radical SAM enzymes involved in the construction of carbon skeletons of cofactors and natural products, Nat. Prod. Rep. 35 (2018) 660–694

K. Yokoyama, Radical breakthroughs in natural product and cofactor biosynthesis, Biochemistry 57 (2018) 390–402.

Kamdar, K. P., Shelton, M. E., & Finnerty, V. (1994). The drosophila molybdenum cofactor gene cinnamon is homologous to three Escherichia coli cofactor proteins and to the rat protein gephyrin. Genetics, 137(3), 791–801. https://doi.org/10.1093/genetics/137.3.791

Kao, A., Chiu, C., Vellucci, D., Yang, Y., Patel, V. R., Guan, S., Randall, A., Baldi, P., Rychnovsky, S. D., & Huang, L. (2011). Development of a Novel Cross-linking Strategy for Fast and Accurate Identification of Cross-linked Peptides of Protein Complexes. Molecular & Cellular Proteomics, 10(1), M110.002170. https://doi.org/10.1074/mcp.m110.002212

Kasaragod, V. B., & Schindelin, H. (2016). Structural Framework for Metal Incorporation during Molybdenum Cofactor Biosynthesis. Structure, 24(5), 782–788. https://doi.org/10.1016/j.str.2016.02.023

Kasaragod, V. B., Hausrat, T. J., Schaefer, N., Kuhn, M., Christensen, N. R., Tessmer, I., Maric, H. M., Madsen, K. L., Sotriffer, C., Villmann, C., Kneussel, M., & Schindelin, H. (2019). Elucidating the Molecular Basis for Inhibitory Neurotransmission Regulation by Artemisinins. Neuron, 101(4), 673-689.e11. https://doi.org/10.1016/j.neuron.2019.01.001

Kaufholdt D, Baillie CK, Bikker R, Burkart V, Dudek CA, von Pein L, Rothkegel M, Mendel RR, Hänsch R. The molybdenum cofactor biosynthesis complex interacts with actin filaments via molybdenum insertase Cnx1 as anchor protein in Arabidopsis thaliana. Plant Sci. 2016 Mar;244:8-18.

Kaufholdt D, Baillie CK, Meyer MH, Schwich OD, Timmerer UL, Tobias L, van Thiel D, Hänsch R, Mendel RR. Identification of a protein-protein interaction network downstream of molybdenum cofactor biosynthesis in Arabidopsis thaliana. J Plant Physiol. 2016 Dec 1;207:42-50.

Keener, J. E., Zhang, G., & Marty, M. T. (2021). Native Mass Spectrometry of Membrane Proteins. Analytical Chemistry, 93(1), 583–597. https://doi.org/10.1021/acs.analchem.0c04342

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2016). Trabajo pr6ctico No 13 . Varianzas en funciyn de variable independiente categyrica. Nature Protocols, 10(6), 845–858. https://doi.org/10.1038/nprot.2015-053

Kim, D. E., Chivian, D., & Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Research, 32(WEB SERVER ISS.), 526–531. https://doi.org/10.1093/nar/gkh468

Kirsch, J., Wolters, I., Triller, A., & Betz, H. (1993). Gephyrin antisense oligonucleotides prevent glycine receptor clustering in spinal neurons. Nature, 366(6457), 745–748. https://doi.org/10.1038/366745a0

Klotz, I. M., Langerman, N. R., & Darnall, D. W. (1970). Quaternary structure of proteins. Annual Review of Biochemistry, 39, 25–62. https://doi.org/10.1146/annurev.bi.39.070170.000325

Kneussel, M., & Betz, H. (2000). Clustering of inhibitory neurotransmitter receptors at developing postsynaptic sites: The membrane activation model. Trends in Neurosciences, 23(9), 429–435. https://doi.org/10.1016/S0166-2236(00)01627-1

Kneussel, M., BrandstΔtter, J. H., Laube, B., Stahl, S., Mьller, U., & Betz, H. (1999). Loss of postsynaptic GABA(A) receptor clustering in gephyrin-deficient mice. Journal of Neuroscience, 19(21), 9289–9297. https://doi.org/10.1523/jneurosci.19-21-09289.1999

Koukos, P. I., Roel-Touris, J., Ambrosetti, F., Geng, C., Schaarschmidt, J., Trellet, M. E., Melquiond, A. S. J., Xue, L. C., Honorato, R. V., Moreira, I., Kurkcuoglu, Z., Vangone, A., & Bonvin, A. M. J. J. (2020). An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. Proteins: Structure, Function and Bioinformatics, 88(8), 1029–1036. https://doi.org/10.1002/prot.25869

Kozakov, D., Hall, D. R., Xia, B., Porter, K. A., Padhorny, D., Yueh, C., Beglov, D., & Vajda, S. (2017). The ClusPro web server for protein-protein docking. Nature Protocols, 12(2), 255–278. https://doi.org/10.1038/nprot.2016.169

Krausze J, Hercher TW, Archna A, Kruse T. The structure of the Moco carrier protein from Rippkaea orientalis. Acta Crystallogr F Struct Biol Commun. 2020 Sep 1;76(Pt 9):453-463.

Krausze, J., Hercher, T. W., Zwerschke, D., Kirk, M. L., Blankenfeldt, W., Mendel, R. R., & Kruse, T. (2018). The functional principle of eukaryotic molybdenum insertases. Biochemical Journal, 475(10), 1739–1753. https://doi.org/10.1042/BCJ20170935

Krausze, J., Probst, C., Curth, U., Reichelt, J., Saha, S., Schafflick, D., Heinz, D. W., Mendel, R. R., & Kruse, T. (2017). Dimerization of the plant molybdenum insertase Cnx1E is required for synthesis of the molybdenum cofactor. Biochemical Journal, 474(1), 163–178. https://doi.org/10.1042/BCJ20160846

Krogan, N. J., Cagney, G., Yu, H., Zhong, G., Guo, X., Ignatchenko, A., Li, J., Pu, S., Datta, N., Tikuisis, A. P., Punna, T., Peregrнn-Alvarez, J. M., Shales, M., Zhang, X., Davey, M., Robinson, M. D., Paccanaro, A., Bray, J. E., Sheung, A., … Greenblatt, J. F. (2006). Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature, 440(7084), 637–643. https://doi.org/10.1038/nature04670

Kruse T, Gehl C, Geisler M, Lehrke M, Ringel P, Hallier S, HΔnsch R, Mendel RR. Identification and biochemical characterization of molybdenum cofactor-binding proteins from Arabidopsis thaliana. J Biol Chem. 2010 Feb 26;285(9):6623-35.

Kuper, J., Llamas, A., Hecht, H. J., Mendel, R. R., & Schwarz, G. (2004). Structure of the molybdopterin-bound Cnx1G domain links molybdenum and copper metabolism. Nature, 430(7001), 803–806. https://doi.org/10.1038/nature02681

Kuper, J., Palmer, T., Mendel, R. R., & Schwarz, G. (2000). Mutations in the molybdenum cofactor biosynthetic protein Cnx1G from Arabidopsis thaliana define functions for molybdopterin binding, molybdenum insertion, and molybdenum cofactor stabilization. Proceedings of the National Academy of Sciences of the United States of America, 97(12), 6475–6480. https://doi.org/10.1073/pnas.110568497

Kuper, J., Winking, J., Hecht, H. J., Mendel, R. R., & Schwarz, G. (2003). The active site of the molybdenum cofactor biosynthetic protein domain Cnx1G. Archives of Biochemistry and Biophysics, 411(1), 36–46. https://doi.org/10.1016/S0003-9861(02)00714-2

Landry, C. R., Levy, E. D., Abd Rabbo, D., Tarassov, K., & Michnick, S. W. (2013). XExtracting insight from noisy cellular networks. Cell, 155(5), 983. https://doi.org/10.1016/j.cell.2013.11.003

Lardi-Studler, B., Smolinsky, B., Petitjean, C. M., Koenig, F., Sidler, C., Meier, J. C., Fritschy, J. M., & Schwarz, G. (2007). Vertebrate-specific sequences in the gephyrin E-domain regulate cytosolic aggregation and postsynaptic clustering. Journal of Cell Science, 120(8), 1371–1382. https://doi.org/10.1242/jcs.003905

Leney, A. C., & Heck, A. J. R. (2017). Native Mass Spectrometry: What is in the Name? Journal of the American Society for Mass Spectrometry, 28(1), 5–13. https://doi.org/10.1007/s13361-016-1545-3

Levy, E. D., De, S., & Teichmann, S. A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proceedings of the National Academy of Sciences of the United States of America, 109(50), 20461–20466. https://doi.org/10.1073/pnas.1209312109

Levy, E. D., Erba, E. B., Robinson, C. V, & Teichmann, S. A. (2009). Europe PMC Funders Group Assembly reflects evolution of protein complexes. Nature, 453(7199), 1262–1265. https://doi.org/10.1038/nature06942.Assembly

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., … Adams, P. D. (2019). Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. Acta Crystallographica Section D: Structural Biology, 75, 861–877. https://doi.org/10.1107/S2059798319011471

Llamas, A., Mendel, R. R., & Schwarz, G. (2004). Synthesis of adenylated molybdopterin: An essential step for molybdenum insertion. Journal of Biological Chemistry, 279(53), 55241–55246. https://doi.org/10.1074/jbc.M409862200

Llamas, A., Otte, T., Multhaup, G., Mendel, R. R., & Schwarz, G. (2006). The mechanism of nucleotide-assisted molybdenum insertion into molybdopterin: A novel route toward metal cofactor assembly. Journal of Biological Chemistry, 281(27), 18343–18350. https://doi.org/10.1074/jbc.M601415200

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., & Marchler-Bauer, A. (2020). CDD/SPARCLE: The conserved

domain database in 2020. Nucleic Acids Research, 48(D1), D265–D268. https://doi.org/10.1093/nar/gkz991

Lunk, H. J., & Hartl, H. (2017). Discovery, properties and applications of molybdenum and its compounds. ChemTexts, 3(3), 1–23. https://doi.org/10.1007/s40828-017-0048-6

Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. Journal of Biological Chemistry, 294(13), 5181–5197. https://doi.org/10.1074/jbc.REV118.005602

M. Christopher, A. M. L. S. (2016). 乳鼠心肌提取 HHS Public Access. Physiology & Behavior, 176(1), 100–106. https://doi.org/10.1002/cpbi.3.Comparative

M. Terao, E. Garattini, M.J. Romao, S. Leimkuhler, Evolution, expression, and substrate specificities of aldehyde oxidase enzymes in eukaryotes, J. Biol. Chem. 295 (2020) 5377–5389.

Ma, B., Elkayam, T., Wolfson, H., & Nussinov, R. (2003). Protein-protein interactions: Structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proceedings of the National Academy of Sciences of the United States of America, 100(10), 5772–5777. https://doi.org/10.1073/pnas.1030237100

Magalon, A., Frixon, C., Pommier, J., Giordano, G., & Blasco, F. (2002). In vivo interactions between gene products involved in the final stages of molybdenum cofactor biosynthesis in Escherichia coli. Journal of Biological Chemistry, 277(50), 48199–48204. https://doi.org/10.1074/jbc.M205806200

Marsh, J. A., & Teichmann, S. A. (2014). Protein Flexibility Facilitates Quaternary Structure Assembly and Evolution. PLoS Biology, 12(5). https://doi.org/10.1371/journal.pbio.1001870

Marsh, J. A., & Teichmann, S. A. (2015). Structure, dynamics, assembly, and evolution of protein complexes. Annual Review of Biochemistry, 84, 551–575. https://doi.org/10.1146/annurev-biochem-060614-034142

Marsh, J. A., Rees, H. A., Ahnert, S. E., & Teichmann, S. A. (2015). Structural and evolutionary versatility in protein complexes with uneven stoichiometry. Nature Communications, 6. https://doi.org/10.1038/ncomms7394

Mashiach, E., Nussinov, R., & Wolfson, H. J. (2010). FiberDock: Flexible induced-fit backbone refinement in molecular docking. Proteins: Structure, Function and Bioinformatics, 78(6), 1503–1519. https://doi.org/10.1002/prot.22668

Mayr, S. J., Mendel, R. R., & Schwarz, G. (2021). Molybdenum cofactor biology, evolution and deficiency. Biochimica et Biophysica Acta - Molecular Cell Research, 1868(1), 118883. https://doi.org/10.1016/j.bbamcr.2020.118883

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40, 658–674

Mendel, R. R. (2013). The molybdenum cofactor. Journal of Biological Chemistry, 288(19), 13165–13172. https://doi.org/10.1074/jbc.R113.455311

Mendel, R. R., & Kruse, T. (2012). Cell biology of molybdenum in plants and humans. Biochimica et Biophysica Acta - Molecular Cell Research, 1823(9), 1568–1579. https://doi.org/10.1016/j.bbamcr.2012.02.007

Mendel, R.R., Alikulov, Z.A., Lvov, N.P., and Mьller, A.J. (1981). Presence of the molybdenum-cofactor in nitrate reductase–deficient mutant cell lines of Nicotiana tabacum. Mol. Gen. Genet. 181 395–399.

Mendel, R.R., and Müller, A.J. (1976). A common genetic determinant of xanthine dehydrogenase and nitrate reductase in Nicotiana tabacum. Biochem. Physiol. Pflanz. 170 538–541.

Mendel, R.R., Marton, L., and Müller, A.J. (1986). Comparative biochemical characterization of nitrate reductase/molybdenum cofactor loci cnxA, cnxB and cnxC of Nicotiana plumbaginifolia. Plant Sci. 43 125–129

Meulen, H. Ter. (1932). Distribution of Molybdenum. Nature, 130(3295), 966.

Meyer, G., Kirsch, J., Betz, H., & Langosch, D. (1995). Identification of a gephyrin binding motif on the glycine receptor β subunit. Neuron, 15(3), 563–572. https://doi.org/10.1016/0896-6273(95)90145-0

Mhaindarkar, D., Gasper, R., Lupilov, N., Hofmann, E., & Leichert, L. I. (2018). Loss of a conserved salt bridge in bacterial glycosyl hydrolase BglM-G1 improves substrate binding in temperate environments. Communications Biology, 1(1), 1–11. https://doi.org/10.1038/s42003-018-0167-7

Ming, W., Laven, J., Krupers, M., Thьne, P. C., Niemantsverdriet, J. W., Brongersma, H. H., & Linde, R. Van Der. (2000). S s p f pmma leis xps.

Mintseris, J., & Gygi, S. P. (2020). High-density chemical cross-linking for modeling protein interactions. Proceedings of the National Academy of Sciences of the United States of America, 117(1), 93–102. https://doi.org/10.1073/pnas.1902931116

Mudd, S. H., Irreverre, F., & Laster, L. (1967). Sulfite oxidase deficiency in man: Demonstration of the enzymatic defect. Science, 156(3782), 1599–1602. https://doi.org/10.1126/science.156.3782.1599

Müller, A.J., and Mendel, R.R. (1989). Biochemical and somatic cell genetics of nitrate reductase in Nicotiana. In Molecular and Genetic Aspects of Nitrate Assimilation, J.L. Wray and J.R. Kinghorn, eds (Oxford: Oxford University Press), pp. 166–185.

Murata, K., & Wolf, M. (2018). Cryo-electron microscopy for structural analysis of dynamic biological macromolecules. Biochimica et Biophysica Acta - General Subjects, 1862(2), 324–334. https://doi.org/10.1016/j.bbagen.2017.07.020

Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P. M. G. E., Grigoras, I. T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E. V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., … Scheres, S. H. W. (2020). Single-particle cryo-EM at atomic resolution. Nature, 587(7832), 152–156. https://doi.org/10.1038/s41586-020-2829-0

Natalello, A., Santambrogio, C., & Grandori, R. (2017). Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? Journal of the American Society for Mass Spectrometry, 28(1), 21–28. https://doi.org/10.1007/s13361-016-1490-1

Nichols, J., & Rajagopalan, K. V. (2002). Escherichia coli MoeA and MogA: Function in metal incorporation step of molybdenum cofactor biosynthesis. Journal of Biological Chemistry, 277(28), 24995–25000. https://doi.org/10.1074/jbc.M203238200

Ohi, M., Li, Y., Cheng, Y., & Walz, T. (2004). Negative staining and image classification - Powerful tools in modern electron microscopy. Biological Procedures Online, 6(1), 23–34. https://doi.org/10.1251/bpo70

Pateman, J. A., & Cove, D. J. (1963). Independently Segregating Genetic Loci Concerned with Nitrate Reductase Activity in Aspergillus nudans. Nature, 198, 262–263.

Perkins, J. R., Diboun, I., Dessailly, B. H., Lees, J. G., & Orengo, C. (2010). Transient Protein-Protein Interactions: Structural, Functional, and Network Properties. Structure, 18(10), 1233–1243. https://doi.org/10.1016/j.str.2010.08.007

Pfeiffer, P., Graham, D., & Betz, H. (1982). Purification by affinity chromatography of the glycine receptor of rat spinal cord. Journal of Biological Chemistry, 257(16), 9389–9393. https://doi.org/10.1016/s0021-9258(18)34082-1

Plate, J., Sassen, W. A., Hassan, A. H., Lehne, F., Kuster, R. W., & Kruse, T. (2019). S-sulfocysteine induces seizure-like behaviors in zebrafish. Frontiers in Pharmacology, 10(APR), 1–6. https://doi.org/10.3389/fphar.2019.00122

Renaud, J. P., Chari, A., Ciferri, C., Liu, W. T., Rйmigy, H. W., Stark, H., & Wiesmann, C. (2018). Cryo-EM in drug discovery: Achievements, limitations and prospects. Nature Reviews Drug Discovery, 17(7), 471–492. https://doi.org/10.1038/nrd.2018.77

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: Advances and challenges. Frontiers in Microbiology, 5(APR), 1–17. https://doi.org/10.3389/fmicb.2014.00172

Roy, A., Kucukural, A., & Zhang, Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. Nature Protocols, 5(4), 725–738. https://doi.org/10.1038/nprot.2010.5

Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V., & Schindelin, H. (2001). Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. Nature Structural Biology, 8(1), 42–46. https://doi.org/10.1038/83034

Rudolph, M. J., Wuebbens, M. M., Rajagopalan, K. V., and Schindelin, H. (2001) Crystal structure of molybdopterin synthase and its evolutionary relationship to ubiquitin activation. Nat. Struct. Biol. 8, 42–46

S. Leimkuhler, The biosynthesis of the molybdenum cofactors in Escherichia coli, Environ. Microbiol. 22 (2020) 2007–2026.

Sander, B., Tria, G., Shkumatov, A. V., Kim, E. Y., Grossmann, J. G., Tessmer, I., Svergun, D. I., & Schindelin, H. (2013). Structural characterization of gephyrin by AFM and SAXS

reveals a mixture of compact and extended states. Acta Crystallographica Section D: Biological Crystallography, 69(10), 2050–2060. https://doi.org/10.1107/S0907444913018714

Schwarz, G., Boxer, D. H., & Mendel, R. R. (1997). Molybdenum cofactor biosynthesis. The plant protein Cnx1 binds molybdopterin with high affinity. Journal of Biological Chemistry, 272(43), 26811–26814. https://doi.org/10.1074/jbc.272.43.26811

Schwarz, G., Boxer, D. H., & Mendel, R. R. (1997). Molybdenum cofactor biosynthesis. The plant protein Cnx1 binds molybdopterin with high affinity. Journal of Biological Chemistry, 272(43), 26811–26814. https://doi.org/10.1074/jbc.272.43.26811

Schwarz, G., Mendel, R. R., & Ribbe, M. W. (2009). Molybdenum cofactors, enzymes and pathways. Nature, 460(7257), 839–847. https://doi.org/10.1038/nature08302

Schwarz, G., Schrader, N., Mendel, R. R., Hecht, H. J., & Schindelin, H. (2001). Crystal structures of human gephyrin and plant Cnx1 G domains: Comparative analysis and functional implications. Journal of Molecular Biology, 312(2), 405–418. https://doi.org/10.1006/jmbi.2001.4952

Schwarz, G., Schulze, J., Bittner, F., Eilers, T., Kuper, J., Bollmann, G., Nerlich, A., Brinkmann, H., & Mendel, R. R. (2000). The molybdenum cofactor biosynthetic protein Cnx1 complements molybdate-repairable mutants, transfers molybdenum to the metal binding pterin, and is associated with the cytoskeleton. Plant Cell, 12(12), 2455–2471. https://doi.org/10.1105/tpc.12.12.2455

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žнdek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7

Services, H. (2013). Toxicological Profile for. Oxidative Medicine and Cellular Longevity, 2013(205), 24. https://doi.org/http://dx.doi.org/10.1155/2013/286524

Simons, K. T., Bonneau, R., Ruczinski, I., & Baker, D. (1999). Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins: Structure, Function and Genetics, 37(SUPPL. 3), 171–176. https://doi.org/10.1002/(SICI)1097-0134(1999)37:3+<171::AID-PROT21>3.0.CO;2-Z

Sinz, A. (2006). Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. Mass Spectrometry Reviews, 25(4), 663–682. https://doi.org/10.1002/mas.20082

Sinz, A., Arlt, C., Chorev, D., & Sharon, M. (2015). Chemical cross-linking and native mass spectrometry: A fruitful combination for structural biology. Protein Science, 24(8), 1193–1209. https://doi.org/10.1002/pro.2696

Smedley, P. L., & Kinniburgh, D. G. (2017). Molybdenum in natural waters: A review of occurrence, distributions and controls. Applied Geochemistry, 84, 387–432. https://doi.org/10.1016/j.apgeochem.2017.05.008

Stallmeyer, B., Schwarz, G., Schulze, J., Nerlich, A., Reiss, J., Kirsch, J., & Mendel, R. R. (1999). The neurotransmitter receptor-anchoring protein gephyrin reconstitutes molybdenum cofactor biosynthesis in bacteria, plants, and mammalian cells. Proceedings of the National Academy of Sciences of the United States of America, 96(4), 1333–1338. https://doi.org/10.1073/pnas.96.4.1333

T. Heidenreich, S. Wollers, R.R. Mendel, F. Bittner, Characterization of the NifSlike domain of ABA3 from Arabidopsis thaliana provides insight into the mechanism of molybdenum cofactor sulfuration, J. Biol. Chem. 280 (2005) 4213–4218.

Tadokoro, T., Kazama, H., Koga, Y., Takano, K., & Kanaya, S. (2013). Investigating the structural dependence of protein stabilization by amino acid substitution. Biochemistry, 52(16), 2839–2847. https://doi.org/10.1021/bi400076f

Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. (2017) STARANISO Cambridge, United Kingdom: Global Phasing Ltd.

Tyagarajan, S. K., & Fritschy, J. M. (2014). Gephyrin: A master regulator of neuronal function? Nature Reviews Neuroscience, 15(3), 141–156. https://doi.org/10.1038/nrn3670

Van Dam, T. J. P., & Snel, B. (2008). Protein complex evolution does not involve extensive network rewiring. PLoS Computational Biology, 4(7). https://doi.org/10.1371/journal.pcbi.1000132

Vimer, S., Ben-Nissan, G., & Sharon, M. (2020). Direct characterization of overproduced proteins by native mass spectrometry. Nature Protocols, 15(2), 236–265. https://doi.org/10.1038/s41596-019-0233-8

Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T and Bricogne G (2011) Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr 67, 293–302

Webb, B., & Sali, A. (2016). Comparative protein structure modeling using MODELLER. Current Protocols in Bioinformatics, 2016(June), 5.6.1-5.6.37. https://doi.org/10.1002/cpbi.3

Wells J.N., Bergendahl L.T., Marsh J.A. (2018) Computational Modelling of Protein Complex Structure and Assembly. In: Marsh J. (eds) Protein Complex Assembly. Methods in Molecular Biology, vol 1764. Humana Press, New York, NY.

Wuebbens, M. M., and Rajagopalan, K. V. (2003) Mechanistic and mutational studies of Escherichia coli molybdopterin synthase clarify the final step of molybdopterin biosynthesis. J. Biol. Chem. 278, 14523–14532

Xiang, S., Nichols, J., Rajagopalan, K. V, Schindelin, H., Brook, S., York, N., & Carolina, N. (2001). The Crystal Structure of Escherichia coli MoeA and Its Relationship to the Multifunctional Protein Gephyrin State University of New York at Stony Brook. Most, 9(01), 299–310.

Yu, C., & Huang, L. (2018). Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. Analytical Chemistry, 90(1), 144–165. https://doi.org/10.1021/acs.analchem.7b04431

Zhang, Y., & Gladyshev, V. N. (2008). Molybdoproteomes and Evolution of Molybdenum Utilization. Journal of Molecular Biology, 379(4), 881–899. https://doi.org/10.1016/j.jmb.2008.03.051

# Supplemental figures and tables

## ✹ Cnx1 XL-MS list

**Table 23. Cross-linking Mass spectrometry of Cnx1 full-length higher molecular band.** Table generated from Mass spectrometry results of the higher molecular weight band of the SDS PAGE generated from crosslinking Cnx1 full-length with DSBU. The table shows the residue number based on Cnx1 with identifier number Q39054. The cross-linking results generated from MeroX are tabulated as the first residue (from residue) to the partner residue (to residue) with domain location, structural availability, type of cross-link and if included in the final modelling step.

| **Higher molecular weight Band on the SDS-PAGE** | | | | | **Cnx1 identifier: Q39054** | | |
|---|---|---|---|---|---|---|---|
| **From residue** | | **To Residue** | | | | | |
| **Domain** | **Number on sequence** | **Domain** | **Number on sequence** | **Structural availability** | **Inter- or intra - domain crosslink** | **Modelled** | **Comment** |
| E-domain | 401 | E-domain | 364 | Yes | Intra | No | |
| E-domain | 307 | E-domain | 158 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| E-domain | 394 | E-domain | 378 | Yes | Intra | No | |
| G-domain | 521 | E-domain | 158 | Yes | Inter | Yes | |
| E-domain | 405 | E-domain | 158 | Yes | Intra | No | |
| E-domain | 307 | E-domain | 31 | Yes | Intra | No | |
| G-domain | 627 | G-domain | 556 | No | Intra | No | Can be fulfilled in the model |
| G-domain | 501 | G-domain | 575 | Yes | Intra | No | Can be fulfilled in the model |
| G-domain | 501 | Linker | 458 | No | Inter | No | Can be fulfilled in the model |
| Linker | 458 | G-domain | 506 | No | Inter | No | Can be fulfilled in the model |
| G-domain | 556 | E-domain | 364 | Yes | Inter | Yes | |
| G-domain | 501 | G-domain | 556 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| E-domain | 307 | E-domain | 283 | Yes | Intra | No | |
| G-domain | 521 | E-domain | 145 | Yes | Inter | Yes | Flexible region |
| E-domain | 394 | G-domain | 556 | Yes | Inter | Yes | |
| G-domain | 501 | E-domain | 158 | Yes | Inter | Yes | |
| G-domain | 556 | E-domain | 378 | Yes | Inter | Yes | |
| E-domain | 307 | E-domain | 285 | Yes | Intra | No | |
| G-domain | 575 | G-domain | 575 | Yes | Intra | No | Trimeric G-domain |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| G-domain | 501 | E-domain | 31 | Yes | Inter | Yes | |
| E-domain | 277 | E-domain | 307 | Yes | Intra | No | |
| Linker | 444 | G-domain | 521 | No | Inter | No | Can be fulfilled in the model |
| E-domain | 401 | E-domain | 378 | Yes | Intra | No | |
| Linker | 444 | E-domain | 378 | No | Inter | No | Can be fulfilled in the model |
| G-domain | 501 | G-domain | 617 | Yes | Intra | No | |
| E-domain | 31 | E-domain | 365 | Yes | Intra | No | |
| E-domain | 31 | G-domain | 630 | No | Inter | No | Can be fulfilled in the model |
| E-domain | 400 | E-domain | 364 | Yes | Intra | No | |
| G-domain | 521 | G-domain | 556 | Yes | Intra | No | |
| Linker | 444 | E-domain | 401 | No | Inter | No | |
| E-domain | 307 | G-domain | 573 | Yes | Inter | Yes | |
| G-domain | 556 | E-domain | 405 | Yes | Inter | Yes | |
| G-domain | 501 | E-domain | 145 | Yes | Inter | Yes | |
| G-domain | 501 | G-domain | 630 | No | Intra | No | Can be fulfilled in the model |
| G-domain | 521 | E-domain | 31 | Yes | Inter | Yes | |
| G-domain | 575 | G-domain | 521 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| E-domain | 233 | E-domain | 307 | Yes | Intra | No | |
| E-domain | 31 | G-domain | 556 | Yes | Inter | Yes | |
| G-domain | 501 | E-domain | 285 | Yes | Inter | Yes | |
| G-domain | 621 | G-domain | 630 | No | Intra | No | Can be fulfilled in the oligomerization model |
| E-domain | 31 | E-domain | 145 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| G-domain | 556 | G-domain | 630 | No | Intra | No | Can be fulfilled in the oligomerization model |
| E-domain | 364 | E-domain | 405 | Yes | Intra | No | |
| E-domain | 307 | G-domain | 506 | Yes | Inter | Yes | |
| G-domain | 575 | E-domain | 364 | Yes | Inter | Yes | |
| Linker | 444 | E-domain | 31 | No | Inter | No | Can be fulfilled in the oligomerization model |
| G-domain | 556 | E-domain | 380 | Yes | Inter | Yes | |
| E-domain | 364 | E-domain | 294 | Yes | Intra | No | |
| G-domain | 556 | G-domain | 556 | Yes | Intra | No | Trimeric G-domain |
| G-domain | 501 | G-domain | 621 | Yes | Intra | No | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| G-domain | 501 | G-domain | 521 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| G-domain | 501 | E-domain | 380 | Yes | Inter | Yes | |
| E-domain | 307 | G-domain | 621 | Yes | Inter | Yes | |
| E-domain | 307 | G-domain | 521 | Yes | Inter | Yes | |
| E-domain | 285 | E-domain | 31 | Yes | Intra | No | |
| G-domain | 498 | E-domain | 145 | Yes | Inter | Yes | |
| E-domain | 394 | G-domain | 521 | Yes | Inter | Yes | |
| E-domain | 307 | G-domain | 556 | Yes | Inter | Yes | |
| G-domain | 575 | G-domain | 573 | Yes | Intra | No | Can be fulfilled in the oligomerization model |
| G-domain | 501 | E-domain | 401 | Yes | Inter | Yes | |
| Linker | 452 | E-domain | 380 | No | Inter | No | Can be fulfilled in the oligomerization model |
| E-domain | 285 | G-domain | 556 | Yes | Inter | Yes | |
| G-domain | 501 | Linker | 444 | No | Inter | No | Can be fulfilled in the oligomerization model |
| E-domain | 307 | E-domain | 401 | Yes | Intra | No | |
| G-domain | 521 | E-domain | 365 | Yes | Inter | Yes | |
| E-domain | 401 | G-domain | 630 | Yes | Inter | No | |
| E-domain | 135 | G-domain | 556 | Yes | Inter | Yes | |
| G-domain | 521 | E-domain | 364 | Yes | Inter | Yes | |
| E-domain | 364 | G-domain | 573 | Yes | Inter | Yes | |
| G-domain | 575 | E-domain | 145 | Yes | Inter | Yes | |
| G-domain | 556 | E-domain | 365 | Yes | Inter | Yes | |
| G-domain | 521 | E-domain | 380 | Yes | Inter | Yes | |
| G-domain | 556 | E-domain | 138 | Yes | Inter | Yes | |
| E-domain | 307 | E-domain | 245 | Yes | Intra | No | |
| G-domain | 556 | E-domain | 145 | Yes | Inter | Yes | |
| E-domain | 31 | Linker | 458 | No | Inter | No | |
| E-domain | 364 | G-domain | 506 | Yes | Inter | Yes | |
| Linker | 444 | E-domain | 380 | No | Inter | No | |
| G-domain | 501 | E-domain | 364 | Yes | Inter | Yes | |
| G-domain | 556 | G-domain | 555 | Yes | Intra | No | |
| E-domain | 401 | G-domain | 556 | Yes | Inter | Yes | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Linker | 444 | G-domain | 556 | No | Inter | No | |
| E-domain | 394 | E-domain | 364 | Yes | Intra | No | |
| G-domain | 617 | G-domain | 499 | Yes | Intra | No | |
| G-domain | 501 | E-domain | 307 | Yes | Inter | Yes | |
| E-domain | 364 | E-domain | 158 | Yes | Intra | No | |
| E-domain | 31 | E-domain | 364 | Yes | Intra | No | |

**Table 24. Cross-linking Mass spectrometry of Cnx1 full-length lower molecular band.** Table generated from Mass spectrometry results of the lower molecular weight band of the SDS PAGE generated from crosslinking Cnx1 full-length with DSBU. The table shows the residue number based on Cnx1 with identifier number Q39054. The cross-linking results generated from MeroX are tabulated as first residue (from residue) to the partner residue (to residue) with domain location, structural availability, type of cross-link and if included in the model.

| **Lower molecular weight Band on the SDS-PAGE** | | | | | **Cnx1 identifier: Q39054** | | |
|---|---|---|---|---|---|---|---|
| **From residue** | | **To Residue** | | | | | |
| **Domain** | **Number on sequence** | **Domain** | **Number on sequence** | **Structural availability** | **Inter- or intra - domain crosslink** | **Modelled** | **Comment** |
| G-domain | 501 | G-domain | 657 | No | Intra | No | |
| E-domain | 401 | E-domain | 364 | Yes | Intra | No | |
| E-domain | 307 | E-domain | 31 | Yes | Intra | No | |
| E-domain | 233 | E-domain | 46 | Yes | Intra | No | |
| E-domain | 31 | E-domain | 364 | Yes | Intra | No | |
| G-domain | 556 | G-domain | 556 | Yes | Intra | No | Trimeric G-domain |
| G-domain | 556 | E-domain | 364 | Yes | Inter | Yes | Included in the model |
| E-domain | 307 | E-domain | 249 | Yes | Intra | No | |
| E-domain | 307 | E-domain | 285 | Yes | Intra | No | |
| G-domain | 521 | G-domain | 556 | Yes | Intra | No | |
| E-domain | 307 | E-domain | 245 | Yes | Intra | No | |
| E-domain | 394 | E-domain | 364 | Yes | Intra | No | |
| E-domain | 233 | E-domain | 307 | Yes | Intra | No | |
| G-domain | 575 | G-domain | 575 | Yes | Intra | No | Trimeric G-domain |
| E-domain | 394 | E-domain | 378 | Yes | Intra | No | |
| G-domain | 627 | G-domain | 556 | No | Intra | No | |
| G-domain | 521 | E-domain | 364 | Yes | Inter | Yes | Included in the model |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| G-domain | 501 | E-domain | 364 | Yes | Inter | Yes | Included in the model |
| G-domain | 501 | G-domain | 617 | Yes | Intra | No | |
| E-domain | 401 | E-domain | 378 | Yes | Intra | No | |
| E-domain | 31 | E-domain | 285 | Yes | Intra | No | |

**Table 25. Residues used for modelling.** Table generated from Mass spectrometry results of the higher molecular weight band of the SDS PAGE generated from crosslinking Cnx1 full-length with DSBU. The table shows the residue number based on Cnx1 with identifier number Q39054. The cross-linking results generated from MeroX are tabulated as the first to second residue used in the final model.

| Inter-domain crosslinks used for Model building | Cnx1 identifier: Q39054 |
|---|---|
| **First residue** | **Second residue** |
| 521 | 158 |
| 501 | 458 |
| 458 | 506 |
| 556 | 364 |
| 521 | 145 |
| 394 | 556 |
| 501 | 158 |
| 556 | 378 |
| 501 | 31 |
| 444 | 521 |
| 444 | 378 |
| 31 | 630 |
| 444 | 401 |
| 307 | 573 |
| 556 | 405 |
| 501 | 145 |
| 521 | 31 |
| 31 | 556 |
| 501 | 285 |
| 307 | 506 |
| 575 | 364 |
| 444 | 31 |
| 556 | 380 |
| 501 | 380 |
| 307 | 621 |
| 307 | 521 |
| 498 | 145 |

| | |
|---|---|
| 394 | 521 |
| 307 | 556 |
| 501 | 401 |
| 452 | 380 |
| 285 | 556 |
| 501 | 444 |
| 521 | 365 |
| 401 | 630 |
| 135 | 556 |
| 521 | 364 |
| 364 | 573 |
| 575 | 145 |
| 556 | 365 |
| 521 | 380 |
| 556 | 138 |
| 556 | 145 |
| 31 | 458 |
| 364 | 506 |
| 444 | 380 |
| 501 | 364 |
| 444 | 556 |
| 401 | 556 |
| 501 | 307 |

# ✹ MeroX software and crosslinking results

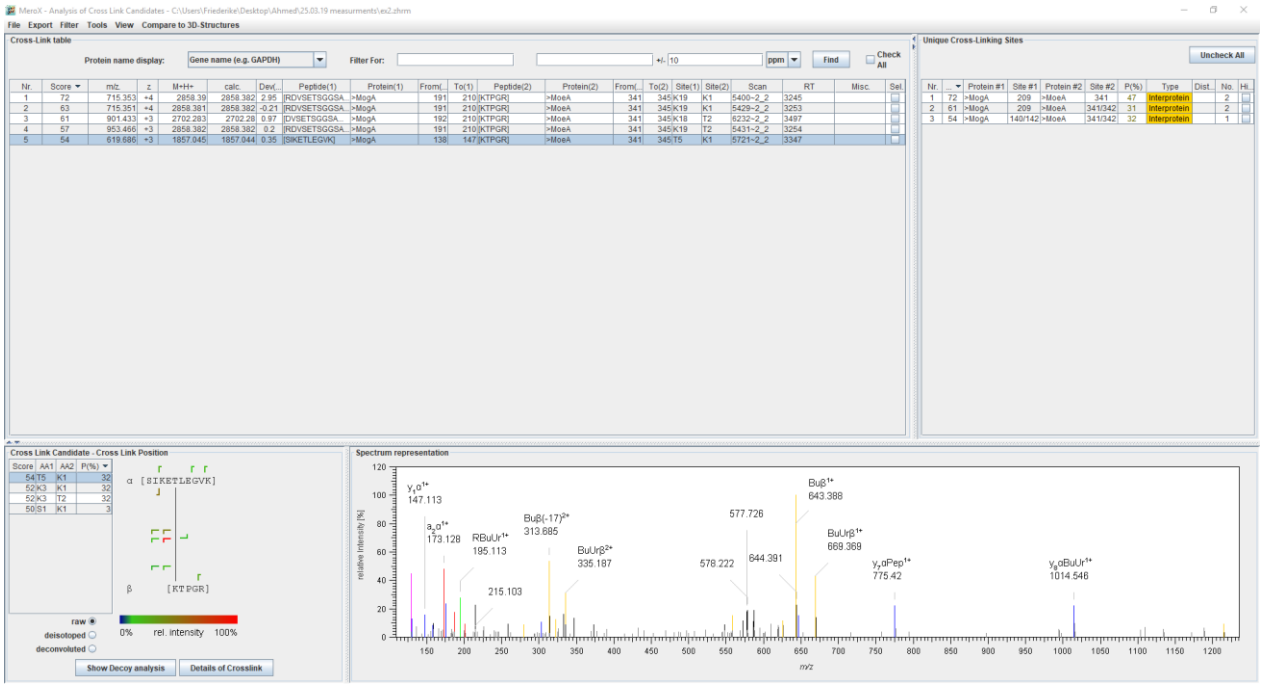



Figure 62. Screenshot of MeroX software results of *in vivo* MoeA-MogA crosslinking.

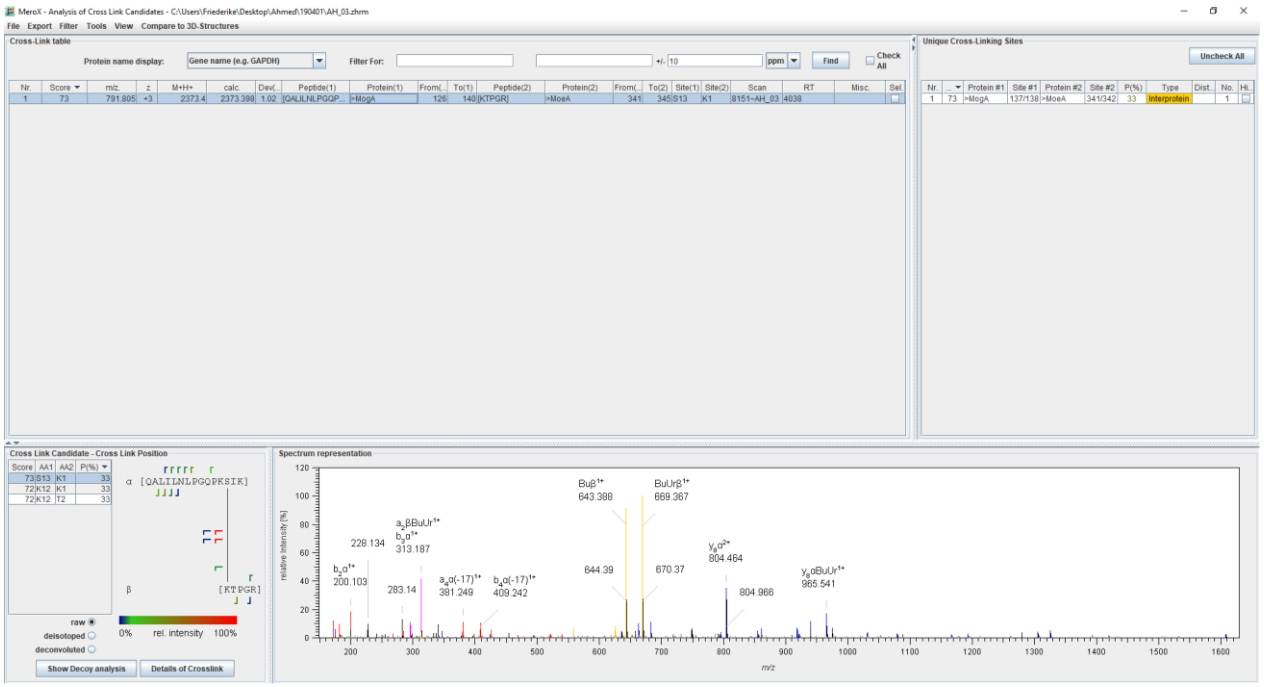



Figure 63. Screenshot of MeroX software results of MoeA-MogA *in vitro* crosslinking.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1 | 315 | >Cnx1 | 556 | >Cnx1 | 556 | 100 | Intraprotein | |
| 2 | 314 | >Cnx1 | 138 | >Cnx1 | 556 | 90 | Intraprotein | |
| 3 | 292 | >Cnx1 | 521 | >Cnx1 | 556 | 100 | Intraprotein | 7.4 ... |
| 4 | 289 | >Cnx1 | 145 | >Cnx1 | 556 | 90 | Intraprotein | |
| 5 | 286 | >Cnx1 | 285 | >Cnx1 | 556 | 83 | Intraprotein | |
| 6 | 286 | >Cnx1 | 31 | >Cnx1 | 556 | 90 | Intraprotein | |
| 7 | 285 | >Cnx1 | 307 | >Cnx1 | 501 | 46 | Intraprotein | |
| 8 | 275 | >Cnx1 | 444 | >Cnx1 | 501 | 38 | Intraprotein | |
| 9 | 274 | >Cnx1 | 556 | >Cnx1 | 627 | 90 | Intraprotein | |
| 10 | 273 | >Cnx1 | 30/31 | >Cnx1 | 501 | 34 | Intraprotein | |
| 11 | 271 | >Cnx1 | 31 | >Cnx1 | 285 | 61 | Intraprotein | 15.... |
| 12 | 268 | >Cnx1 | 621 | >Cnx1 | 630 | 100 | Intraprotein | |
| 13 | 262 | >Cnx1 | 499/501 | >Cnx1 | 617 | 38 | Intraprotein | 6.8 ... |
| 14 | 260 | >Cnx1 | 555 | >Cnx1 | 556 | 71 | Intraprotein | 3.8 ... |
| 15 | 260 | KRT38 | 256 | >Cnx1 | 556 | 100 | Interprotein | |
| 16 | 258 | >Cnx1 | 378 | >Cnx1 | 394 | 47 | Intraprotein | 11.... |
| 17 | 256 | >Cnx1 | 400/401 | >Cnx1 | 556 | 42 | Intraprotein | |
| 18 | 255 | >Cnx1 | 501 | >Cnx1 | 621 | 52 | Intraprotein | 12.... |
| 19 | 254 | >Cnx1 | 245 | >Cnx1 | 307 | 73 | Intraprotein | |
| 20 | 251 | >Cnx1 | 31 | >Cnx1 | 307 | 61 | Intraprotein | |
| 21 | 238 | >Cnx1 | 378 | >Cnx1 | 556 | 100 | Intraprotein | |
| 22 | 236 | >Cnx1 | 307 | >Cnx1 | 556 | 71 | Intraprotein | |
| 23 | 235 | >Cnx1 | 378 | >Cnx1 | 444 | 61 | Intraprotein | |
| 24 | 233 | HARS | 22 | >Cnx1 | 521 | 100 | Interprotein | |
| 25 | 225 | >Cnx1 | 556 | >Cnx1 | 630 | 100 | Intraprotein | |
| 26 | 218 | >Cnx1 | 145 | >Cnx1 | 521 | 90 | Intraprotein | |
| 27 | 218 | >Cnx1 | 501 | >Cnx1 | 617 | 52 | Intraprotein | 6.8 ... |
| 28 | 215 | >Cnx1 | 277 | >Cnx1 | 307 | 73 | Intraprotein | |
| 29 | 213 | >Cnx1 | 158 | >Cnx1 | 521 | 61 | Intraprotein | |
| 30 | 213 | >Cnx1 | 30/31 | >Cnx1 | 444 | 39 | Intraprotein | |
| 31 | 212 | >Cnx1 | 364 | >Cnx1 | 556 | 90 | Intraprotein | |
| 32 | 210 | >Cnx1 | 233 | >Cnx1 | 307 | 83 | Intraprotein | |

**Figure 64. Screenshot of Cnx1 cross-linking results on MeroX.** All cross-links shown are intra protein but inter-domain (within Cnx1 but from E- to G- domain or linker.

# ✹ Insilico digestion of MoeA and/or MogA with Trypsin and GluC

## MoeA trypsin

| mass | position | #MC | modifications | peptide sequence |
|---|---|---|---|---|
| 3565.7016 | 356-388 | 0 | | NADGELEVTTTGHQGSHIFS SFSLGNCFIVLER |
| 3264.7125 | 289-318 | 0 | | LSNSWFCGLPGNPVSATLTF YQLVQPLLAK |
| 3019.3627 | 98-125 | 0 | | IMTGAPVPEGCEAVVMQEQT EQMDNGVR |
| 2904.4373 | 180-206 | 0 | | VALFSTGDELQLPGQPLGDG QIYDTNR |
| 2786.3366 | 235-262 | 0 | | AAFIEADSQADVVISSGGVS VGEADYTK |
| 2708.3235 | 39-64 | 0 | | ILASDVVSPLDVPGFDNSAM DGYAVR |
| 2334.3035 | 207-227 | 0 | | LAVHLMLEQLGCEVINLGII R |
| 2262.1076 | 391-411 | 0 | | GNVEVGEWVEVEPFNALFGG L |
| 2260.0658 | 1-20 | 0 | | MEFTTGLMSLDTALNEMLSR |
| 2162.2794 | 156-176 | 0 | | LTTAELPVIASLGIAEVPVI R |
| 1976.9071 | 80-97 | 0 | | SFAGQPYHGEWPAGTCIR |
| 1973.0524 | 21-38 | 0 | | VTPLTAQETLPLVQCFGR |
| 1548.8308 | 263-275 | 0 | | TILEELGEIAFWK |
| 1546.7859 | 140-155 | 0 | | GEDISAGAVVFPAGTR |
| 1436.8107 | 65-79 | 0 | | LADIASGQPLPVAGK |
| 1373.8303 | 276-288 | 0 | | LAIKPGKPFAFGK |
| | | | | |
| 1143.6116 | 319-330 | 0 | | LSGNTASGLPAR |
| 823.4057 | 228-234 | 0 | | DDPHALR |
| 722.3831 | 126-131 | 0 | | FTAEVR |
| 678.3569 | 346-350 | 0 | | LDFQR |
| 674.3580 | 132-137 | 0 | | SGQNIR |
| 572.3514 | 351-355 | 0 | | GVLQR |

## MoeA GluC

| mass | position | #MC | modifications | peptide sequence |
|---|---|---|---|---|
| 8352.6378 | 271-347 | 0 | | IAFWKLAIKPGKPFAFGKLS NSWFCGLPGNPVSATLTFYQ LVQPLLAKLSGNTASGLPAR QRVRTASRLKKTPGRLD |
| 2681.3028 | 363-387 | 0 | | VTTTGHQGSHIFSSFSLGNC FIVLE |
| 2211.1192 | 68-89 | 0 | | IASGQPLPVAGKSFAGQPYH GE |
| 1886.0857 | 172-188 | 0 | | VPVIRKVRVALFSTGDE |
| 1798.8978 | 90-106 | 0 | | WPAGTCIRIMTGAPVPE |
| 1760.9541 | 143-160 | 0 | | ISAGAVVFPAGTRLTTAE |
| 1632.8777 | 29-43 | 0 | | TLPLVQCFGRILASD |
| 1427.7938 | 130-141 | 0 | | VRSGQNIRRRGE |
| 1345.7144 | 17-28 | 0 | | MLSRVTPLTAQE |
| 1303.6865 | 348-358 | 0 | | FQRGVLQRNAD |
| 1296.7092 | 204-214 | 0 | | TNRLAVHLMLE |
| 1239.6480 | 229-239 | 0 | | DPHALRAAFIE |
| 1163.6095 | 401-411 | 0 | | VEPFNALFGGL |
| 1089.5786 | 246-257 | 0 | | VVISSGGVSVGE |
| 1082.6455 | 161-171 | 0 | | LPVIASLGIAE |
| 1037.5625 | 189-198 | 0 | | LQLPGQPLGD |
| 1012.6149 | 220-228 | 0 | | VINLGIIRD |
| 996.5248 | 260-267 | 0 | | YTKTILEE |
| 984.4706 | 3-11 | 0 | | FTTGLMSLD |
| 893.4475 | 122-129 | 0 | | NGVRFTAE |
| 864.4574 | 60-67 | 0 | | GYAVRLAD |
| 676.3334 | 110-115 | 0 | | AVVMQE |
| 629.3504 | 44-49 | 0 | | VVSPLD |
| 595.2722 | 199-203 | 0 | | GQIYD |
| 574.2943 | 390-394 | 0 | | RGNVE |
| 549.2337 | 215-219 | 0 | | QLGCE |
| 547.2722 | 12-16 | 0 | | TALNE |
| 537.1973 | 55-59 | 0 | | NSAMD |
| 534.2558 | 50-54 | 0 | | VPGFD |

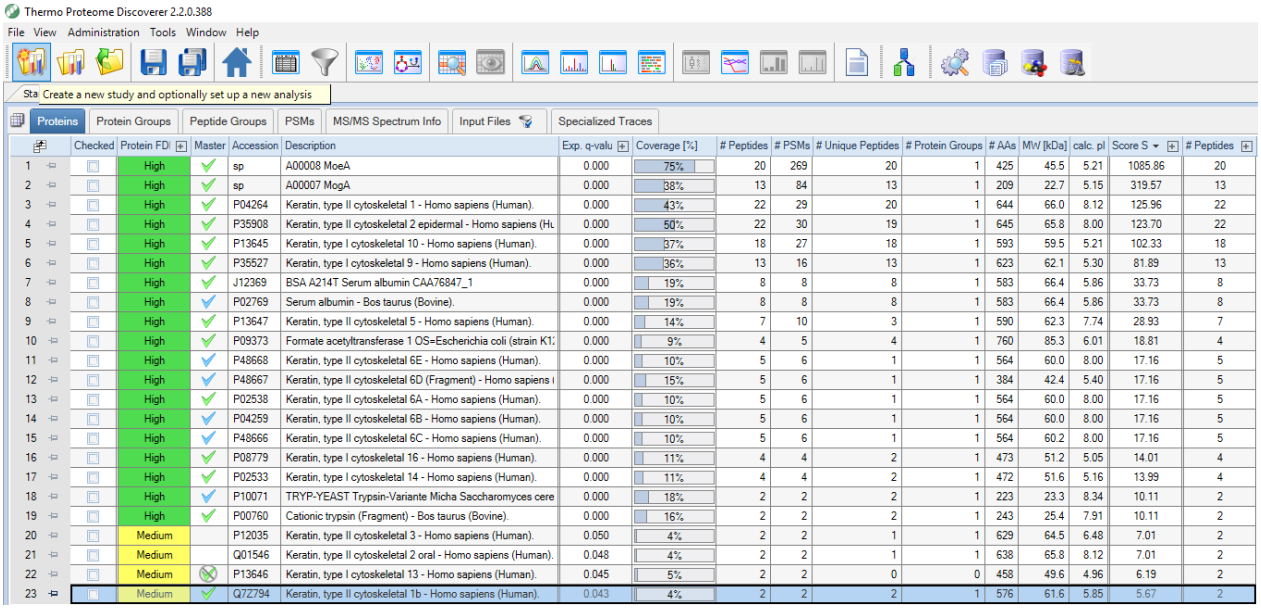
## MogA trypsine

| mass | position | #MC | modifications | peptide sequence |
|---|---|---|---|---|
| 4285.2001 | 148-187 | 0 | | DAEGNVVVHGIFASVPYCIQ LLEGPYVETAPEVVAAFRPK |
| 3752.8543 | 46-80 | 0 | | LIPDEQAIIEQTLCELVDEM SCHLVLTTGGTGPAR |
| 2474.2813 | 24-45 | 0 | | GIPALEEWLTSALTTPFELE TR |
| 1581.9111 | 105-118 | 0 | | QISLHFVPTAILSR |
| 1343.6801 | 82-94 | 0 | | DVTPDATLAVADR |
| 1291.7732 | 126-137 | 0 | | QALILNLPGQPK |
| 1181.5078 | 95-104 | 0 | | EMPGFGEQMR |
| 959.5520 | 6-14 | 0 | | IGLVSISDR |
| 954.4527 | 15-23 | 0 | | ASSGVYQDK |
| 775.4196 | 141-147 | 0 | | ETLEGVK |
| 671.4199 | 119-124 | 0 | | QVGVIR |
| 634.3341 | 1-5 | 0 | | MNTLR |

## MogA GluC

| mass | position | #MC | modifications | peptide sequence |
|---|---|---|---|---|
| 4507.6173 | 102-141 | 0 | | QMRQISLHFVPTAILSRQVG VIRKQALILNLPGQPKSIKE |
| 2158.1365 | 151-170 | 0 | | GNVVVHGIFASVPYCIQLLE |
| 1871.9214 | 65-82 | 0 | | MSCHLVLTTGGTGPARRD |
| 1472.8444 | 180-192 | 0 | | VVAAFRPKSARRD |
| 1418.7671 | 1-13 | 0 | | MNTLRIGLVSISD |
| 1265.6412 | 31-41 | 0 | | WLTSALTTPFE |
| 982.4588 | 14-22 | 0 | | RASSGVYQD |
| 856.4774 | 23-30 | 0 | | KGIPALEE |
| 843.4570 | 44-50 | 0 | | TRLIPDE |
| 660.3562 | 87-93 | 0 | | ATLAVAD |
| 637.2650 | 96-101 | 0 | | MPGFGE |
| 593.2599 | 56-60 | 0 | | QTLCE |
| 573.3242 | 51-55 | 0 | | QAIIE |
| 564.2664 | 171-175 | 0 | | GPYVE |

# ✹ Screenshot showing MoeA-MogA cross-linking with other proteins and keratin




| | Checked | Protein FDI | Master | Accession | Description | Exp. q-valu | Coverage [%] | # Peptides | # PSMs | # Unique Peptides | # Protein Groups | # AAs | MW [kDa] | calc. pI | Score S | # Peptides |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | High | ✓ | sp | A00008 MoeA | 0.000 | 75% | 20 | 269 | 20 | 1 | 425 | 45.5 | 5.21 | 1085.86 | 20 |
| 2 | | High | ✓ | sp | A00007 MogA | 0.000 | 38% | 13 | 84 | 13 | 1 | 209 | 22.7 | 5.15 | 319.57 | 13 |
| 3 | | High | ✓ | P04264 | Keratin, type II cytoskeletal 1 - Homo sapiens (Human). | 0.000 | 43% | 22 | 29 | 20 | 1 | 644 | 66.0 | 8.12 | 125.96 | 22 |
| 4 | | High | ✓ | P35908 | Keratin, type II cytoskeletal 2 epidermal - Homo sapiens (Hu | 0.000 | 50% | 22 | 30 | 19 | 1 | 645 | 65.8 | 8.00 | 123.70 | 22 |
| 5 | | High | ✓ | P13645 | Keratin, type I cytoskeletal 10 - Homo sapiens (Human). | 0.000 | 37% | 18 | 27 | 18 | 1 | 593 | 59.5 | 5.21 | 102.33 | 18 |
| 6 | | High | ✓ | P35527 | Keratin, type I cytoskeletal 9 - Homo sapiens (Human). | 0.000 | 36% | 13 | 16 | 13 | 1 | 623 | 62.1 | 5.30 | 81.89 | 13 |
| 7 | | High | ✓ | J12369 | BSA A214T Serum albumin CAA76847_1 | 0.000 | 19% | 8 | 8 | 8 | 1 | 583 | 66.4 | 5.86 | 33.73 | 8 |
| 8 | | High | ✓ | P02769 | Serum albumin - Bos taurus (Bovine). | 0.000 | 19% | 8 | 8 | 8 | 1 | 583 | 66.4 | 5.86 | 33.73 | 8 |
| 9 | | High | ✓ | P13647 | Keratin, type II cytoskeletal 5 - Homo sapiens (Human). | 0.000 | 14% | 7 | 10 | 3 | 1 | 590 | 62.3 | 7.74 | 28.93 | 7 |
| 10 | | High | ✓ | P09373 | Formate acetyltransferase 1 OS=Escherichia coli (strain K1 | 0.000 | 9% | 4 | 5 | 4 | 1 | 760 | 85.3 | 6.01 | 18.81 | 4 |
| 11 | | High | ✓ | P48668 | Keratin, type II cytoskeletal 6E - Homo sapiens (Human). | 0.000 | 10% | 5 | 6 | 1 | 1 | 564 | 60.0 | 8.00 | 17.16 | 5 |
| 12 | | High | ✓ | P48667 | Keratin, type II cytoskeletal 6D (Fragment) - Homo sapiens ( | 0.000 | 15% | 5 | 6 | 1 | 1 | 384 | 42.4 | 5.40 | 17.16 | 5 |
| 13 | | High | ✓ | P02538 | Keratin, type II cytoskeletal 6A - Homo sapiens (Human). | 0.000 | 10% | 5 | 6 | 1 | 1 | 564 | 60.0 | 8.00 | 17.16 | 5 |
| 14 | | High | ✓ | P04259 | Keratin, type II cytoskeletal 6B - Homo sapiens (Human). | 0.000 | 10% | 5 | 6 | 1 | 1 | 564 | 60.0 | 8.00 | 17.16 | 5 |
| 15 | | High | ✓ | P48666 | Keratin, type II cytoskeletal 6C - Homo sapiens (Human). | 0.000 | 10% | 5 | 6 | 1 | 1 | 564 | 60.2 | 8.00 | 17.16 | 5 |
| 16 | | High | ✓ | P08779 | Keratin, type I cytoskeletal 16 - Homo sapiens (Human). | 0.000 | 11% | 4 | 4 | 2 | 1 | 473 | 51.2 | 5.05 | 14.01 | 4 |
| 17 | | High | ✓ | P02533 | Keratin, type I cytoskeletal 14 - Homo sapiens (Human). | 0.000 | 11% | 4 | 4 | 2 | 1 | 472 | 51.6 | 5.16 | 13.99 | 4 |
| 18 | | High | ✓ | P10071 | TRYP-YEAST Trypsin-Variante Micha Saccharomyces cere | 0.000 | 18% | 2 | 2 | 2 | 1 | 223 | 23.3 | 8.34 | 10.11 | 2 |
| 19 | | High | ✓ | P00760 | Cationic trypsin (Fragment) - Bos taurus (Bovine). | 0.000 | 16% | 2 | 2 | 2 | 1 | 243 | 25.4 | 7.91 | 10.11 | 2 |
| 20 | | Medium | | P12035 | Keratin, type II cytoskeletal 3 - Homo sapiens (Human). | 0.050 | 4% | 2 | 2 | 1 | 1 | 629 | 64.5 | 6.48 | 7.01 | 2 |
| 21 | | Medium | | Q01546 | Keratin, type II cytoskeletal 2 oral - Homo sapiens (Human). | 0.048 | 4% | 2 | 2 | 1 | 1 | 638 | 65.8 | 8.12 | 7.01 | 2 |
| 22 | | Medium | ⊗ | P13646 | Keratin, type I cytoskeletal 13 - Homo sapiens (Human). | 0.045 | 5% | 2 | 2 | 0 | 0 | 458 | 49.6 | 4.96 | 6.19 | 2 |
| 23 | | Medium | ✓ | Q7Z794 | Keratin, type II cytoskeletal 1b - Homo sapiens (Human). | 0.043 | 4% | 2 | 2 | 2 | 1 | 576 | 61.6 | 5.85 | 5.67 | 2 |

**Figure 65. Screenshot showing MoeA-MogA cross-linking with other proteins and keratin.** Proteome discoverer screenshot of MoeA-MogA cross-linking showing high keratin contamination.

## ✹ Western blot of MoeA-MogA with different cross-linkers

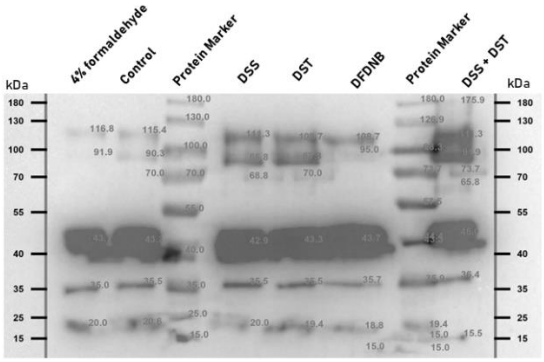



**Figure 66. Western blot of MoeA-MogA crosslinking with different crosslinkers.** The blot generated from MoeA-MogA 5-fold excess crosslinking and visualized using Anti-strep tag antibodies. In the case of DSS+DST, a 10-fold excess was added. The blot shows high molecular weight band in the case of DSS+DST.

## ✹ Native MS screenshot

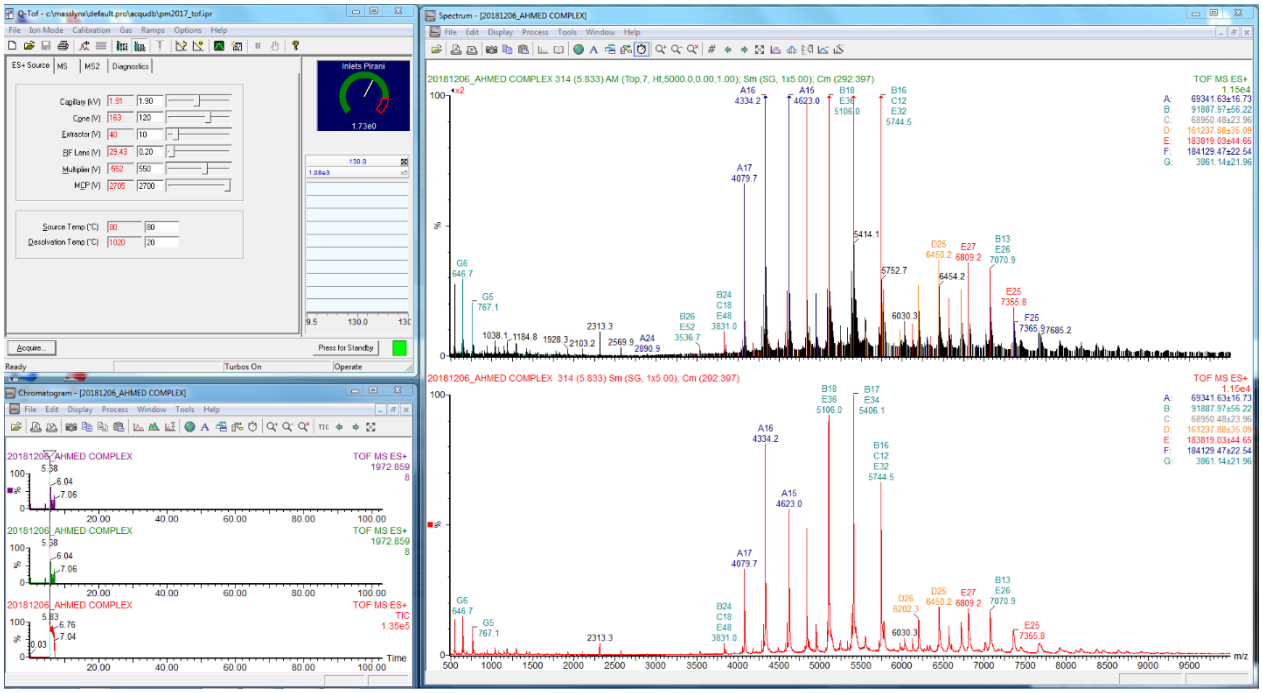



**Figure 67. Screenshot of Native MS results of MoeA-MogA *in vitro* mix.**

## ✹ E. coli crosslinking with DSSO and DSBU

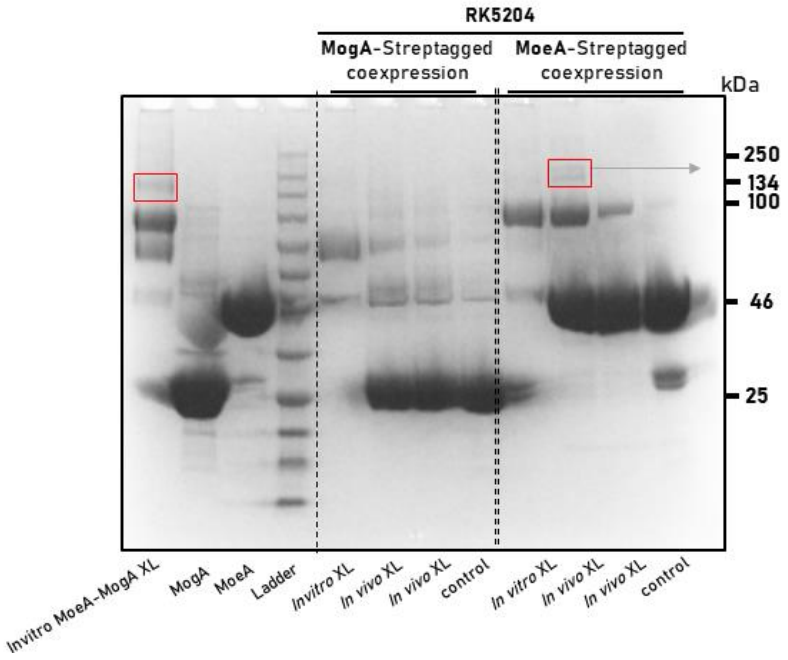



**Figure 68. SDS PAGE of *in vivo* and *in vitro* MoeA-MogA crosslinking with DSBU.** All crosslinking was done using DSBU in 100-fold excess to the protein. For *in vivo* crosslinking, RK5204 was used. All *in vivo* crosslinking was done with co-expression of MoeA and MogA. This was done with either MoeA or MogA C-terminus strep tagged. High molecular weight bands used for LC-MS/MS are highlighted in a red rectangle.

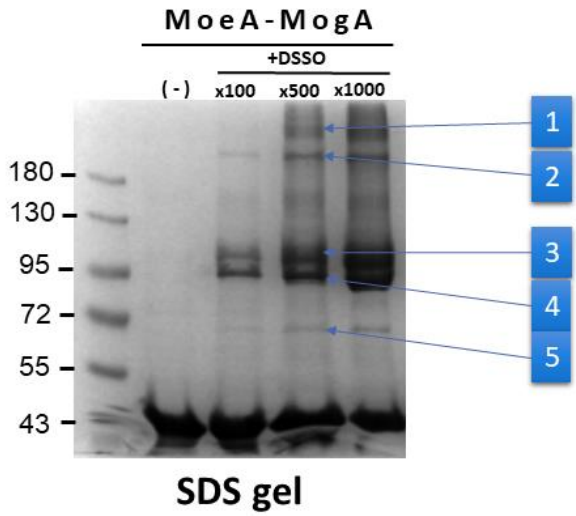



**Figure 69. SDS PAGE of in vivo MoeA-MogA crosslinking with DSSO.** All crosslinking was done using DSSO in 100-fold, 500-fold and 1000-fold molar excess to the proteins. For *in vivo* crosslinking, RK5204 was used. All *in vivo* crosslinking was done with co-expression of MoeA and MogA. This was done with either MoeA or MogA C-terminus strep tagged. Bands used for LC-MS/MS are marked in blue from 1 to 5. Table 26 are the results of this SDS PAGE.

# ✹ MS results of E. coli crosslinking with DSSO

**Table 26. MS results of MoeA-MogA crosslinking with DSSO.** The table generated from the SDS PAGE shown in figure 65. The table shows coverage of each band excised from the gel. Control band was used as a control from the lower molecular weight bands, showing only MoeA. All high molecular weight bands showed both MoeA and MogA. (These results are generated with the help of Prof. Lothar Jänsch)

| Accession | Description | Coverage [%] | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 | Control | Score Sequest HAT | # Unique Peptides | # PSMs | # Peptides | MW [kDa] | Oxidation (M) | Oxidation (M) Positions |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MoeA | MoeA | 79 | 1.78E+09 | 1.80E+09 | 3.24E+09 | 7.89E+08 | 6.61E+07 | 1.34E+09 | 1.40E+09 | 21402.05 | 21 | 7156 | 21 | 44 | 8 | Oxidation [M1; M8; M17; M58; M99; M113; M120; M212] |
| P0CE47 | Elongation factor Tu 1 OS=Escherichia coli (strain K12) OX=83333 GN=tufA PE=1 SV=1 | 76 | 6.89E+07 | 8.11E+07 | 3.53E+07 | 7.36E+06 | 6.07E+05 | 1.62E+06 | 2.94E+05 | 961.05 | 21 | 352 | 21 | 43.3 | 8 | Oxidation [M92; M99; M113; M261; M350; M352; M359; M369] |
| P0A9Q7 | Aldehyde-alcohol dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=adhE PE=1 SV=2 | 29 | 3.43E+06 | 4.76E+05 | 1.79E+07 | 6.97E+04 | | 1.56E+04 | | 199.13 | 20 | 70 | 20 | 96.1 | 5 | Oxidation [M52; M59; M236; M343; M592] |
| P0A6M8 | Elongation factor G OS=Escherichia coli (strain K12) OX=83333 GN=fusA PE=1 SV=2 | 39 | 1.65E+07 | 7.12E+06 | 1.04E+06 | 1.41E+06 | | | | 272.45 | 21 | 101 | 21 | 77.5 | 5 | Oxidation [M147; M366; M409; M431; M617] |
| P0C8J8 | D-tagatose-1,6-bisphosphate aldolase subunit GatZ OS=Escherichia coli (strain K12) OX=83333 GN=gatZ PE=1 SV=1 | 35 | 3.03E+07 | 3.00E+07 | 1.73E+07 | 1.86E+06 | 1.02E+06 | 4.87E+05 | | 435.52 | 11 | 188 | 11 | 47.1 | 4 | Oxidation [M59; M127; M252; M404] |
| P09373 | Formate acetyltransferase 1 OS=Escherichia coli (strain K12) OX=83333 GN=pflB PE=1 SV=2 | 24 | 9.64E+06 | 3.19E+06 | 2.68E+06 | 1.40E+06 | 2.59E+05 | | | 160.71 | 13 | 71 | 14 | 85.3 | 6 | Oxidation [M194; M430; M477; M512; M523; M637] |
| P0AF03 | Molybdopterin adenylyltransferase OS=Escherichia coli (strain K12) OX=83333 GN=mog PE=1 SV=1 | 41 | 7.54E+07 | 4.82E+07 | 9.38E+06 | 3.28E+06 | 5.73E+07 | 1.54E+06 | | 618.26 | 4 | 277 | 11 | 21.2 | 2 | Oxidation [M96; M103] |
| P08839 | Phosphoenolpyruvate-protein phosphotransferase OS=Escherichia coli (strain K12) OX=83333 GN=ptsI PE=1 SV=1 | 29 | 1.36E+06 | 5.35E+05 | 1.33E+05 | | 1.56E+06 | | | 150.82 | 13 | 54 | 13 | 63.5 | 5 | Oxidation [M1; M302; M334; M345; M562] |
| P36683 | Aconitate hydratase B OS=Escherichia coli (strain K12) OX=83333 GN=acnB PE=1 SV=3 | 18 | 1.30E+07 | 5.55E+06 | 5.18E+06 | 7.67E+05 | 6.44E+05 | 2.23E+04 | 6.25E+04 | 100.38 | 12 | 43 | 12 | 93.4 | 4 | Oxidation [M408; M420; M486; M743] |
| P0AFF6 | Transcription termination/antitermination protein NusA OS=Escherichia coli (strain K12) OX=83333 GN=nusA PE=1 SV=1 | 23 | 6.53E+05 | 4.42E+04 | 1.15E+05 | | 8.69E+05 | | | 74.16 | 11 | 29 | 11 | 54.8 | 3 | Oxidation [M125; M254; M483] |
| P16659 | Proline--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=proS PE=1 SV=4 | 21 | 1.56E+05 | | 4.40E+04 | 1.17E+04 | 7.30E+05 | | 1.02E+04 | 117.03 | 10 | 36 | 10 | 63.7 | 3 | Oxidation [M188; M252; M343] |
| P0AB53 | Tryptophanase OS=Escherichia coli (strain K12) OX=83333 GN=tnaA PE=1 SV=1 | 26 | 8.45E+06 | 5.57E+06 | 4.55E+06 | 1.26E+06 | 5.43E+04 | 4.52E+05 | 1.45E+05 | 212.21 | 10 | 77 | 10 | 52.7 | 6 | Oxidation [M213; M226; M263; M266; M315; M432] |
| MogA | MogA | 31 | 1.35E+06 | 6.77E+05 | 8.19E+04 | 1.74E+04 | 6.35E+05 | 2.80E+04 | | 374.73 | 1 | 172 | 8 | 21.4 | 2 | Oxidation [M96; M103] |
| P0A799 | Phosphoglycerate kinase OS=Escherichia coli (strain K12) OX=83333 GN=pgk PE=1 SV=2 | 35 | 9.58E+06 | 8.44E+06 | 6.16E+06 | 7.33E+05 | 7.44E+04 | 2.76E+04 | | 118.58 | 10 | 52 | 10 | 41.1 | 3 | Oxidation [M6; M138; M380] |
| P0A9B2 | Glyceraldehyde-3-phosphate dehydrogenase A OS=Escherichia coli (strain K12) OX=83333 GN=gapA PE=1 SV=2 | 30 | 1.77E+07 | 8.56E+06 | 4.41E+06 | 1.49E+06 | 3.94E+05 | 1.95E+05 | | 171.1 | 9 | 94 | 9 | 35.5 | 1 | Oxidation [M229] |
| P0A6P9 | Enolase OS=Escherichia coli (strain K12) OX=83333 GN=eno PE=1 SV=2 | 22 | 5.55E+06 | 3.25E+06 | 7.52E+05 | 2.83E+05 | | 3.59E+05 | 1.88E+05 | 124.49 | 7 | 44 | 7 | 45.6 | 2 | Oxidation [M94; M245] |
| P00957 | Alanine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=alaS PE=1 SV=2 | 12 | 3.84E+05 | 8.89E+05 | 1.88E+05 | 3.38E+05 | | | | 38.44 | 8 | 14 | 8 | 96 | 3 | Oxidation [M449; M604; M631] |
| P0A6P1 | Elongation factor Ts OS=Escherichia coli (strain K12) OX=83333 GN=tsf PE=1 SV=2 | 29 | 6.92E+06 | 1.83E+06 | 8.12E+05 | 3.46E+05 | | | | 34.46 | 7 | 15 | 7 | 30.4 | 1 | Oxidation [M236] |
| P0AB85 | Lysine--tRNA ligase, heat inducible OS=Escherichia coli (strain K12) OX=83333 GN=lysU PE=1 SV=2 | 18 | 1.08E+06 | 2.85E+05 | 4.07E+05 | 2.33E+04 | 5.61E+05 | | | 35.06 | 9 | 20 | 9 | 57.8 | 0 | |
| P03023 | Lactose operon repressor OS=Escherichia coli (strain K12) OX=83333 GN=lacI PE=1 SV=3 | 28 | 7.43E+06 | 4.72E+06 | 1.26E+06 | 5.73E+05 | 5.51E+04 | 1.95E+05 | | 107.01 | 8 | 49 | 8 | 38.6 | 1 | Oxidation [M347] |
| P07395 | Phenylalanine--tRNA ligase beta subunit OS=Escherichia coli (strain K12) OX=83333 GN=pheT PE=1 SV=2 | 10 | 7.13E+05 | 7.30E+04 | 6.37E+05 | 6.66E+05 | 9.25E+03 | 1.24E+05 | | 26.82 | 7 | 13 | 7 | 87.3 | 0 | |
| P0A9S3 | Galactitol 1-phosphate 5-dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=gatD PE=1 SV=1 | 19 | 4.22E+06 | 1.57E+06 | 1.41E+06 | 1.00E+05 | | | | 25.84 | 6 | 14 | 6 | 37.4 | 0 | |
| P00962 | Glutamine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=glnS PE=1 SV=3 | 12 | 1.73E+05 | 1.50E+05 | 5.14E+04 | | 5.13E+05 | | | 31.78 | 6 | 12 | 6 | 63.4 | 4 | Oxidation [M179; M186; M269; M344] |
| P21889 | Aspartate--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=aspS PE=1 SV=1 | 11 | 3.82E+05 | 3.35E+04 | 4.90E+04 | | 3.01E+05 | | | 37.65 | 5 | 15 | 5 | 65.9 | 2 | Oxidation [M270; M290] |
| P68066 | Autonomous glycyl radical cofactor OS=Escherichia coli (strain K12) OX=83333 GN=grcA PE=1 SV=1 | 47 | 3.11E+06 | 1.15E+06 | 1.50E+06 | 2.75E+05 | 2.84E+05 | | | 40.48 | 5 | 29 | 6 | 14.3 | 1 | Oxidation [M1] |
| P0AC41 | Succinate dehydrogenase flavoprotein subunit OS=Escherichia coli (strain K12) OX=83333 GN=sdhA PE=1 SV=1 | 7 | 1.15E+05 | 5.24E+04 | | | 3.41E+05 | | | 24.16 | 4 | 10 | 4 | 64.4 | 0 | |
| P0AB71 | Fructose-bisphosphate aldolase class 2 OS=Escherichia coli (strain K12) OX=83333 GN=fbaA PE=1 SV=2 | 9 | 1.93E+06 | 7.08E+05 | 4.38E+05 | 1.68E+05 | | | | 36.98 | 2 | 11 | 2 | 39.1 | 0 | |
| P0A7D4 | Adenylosuccinate synthetase OS=Escherichia coli (strain K12) OX=83333 GN=purA PE=1 SV=2 | 13 | 1.47E+06 | 1.67E+06 | 8.06E+05 | 1.50E+05 | 8.70E+03 | 4.52E+05 | 1.22E+05 | 35.6 | 5 | 20 | 5 | 47.3 | 0 | |
| P0AFG8 | Pyruvate dehydrogenase E1 component OS=Escherichia coli (strain K12) OX=83333 GN=aceE PE=1 SV=2 | 5 | 1.47E+06 | 1.71E+05 | 3.08E+06 | 5.98E+05 | | | | 37.24 | 5 | 22 | 5 | 99.6 | 0 | |
| P33195 | Glycine dehydrogenase (decarboxylating) OS=Escherichia coli (strain K12) OX=83333 GN=gcvP PE=1 SV=3 | 5 | 7.72E+05 | 5.35E+04 | 3.15E+05 | 3.91E+05 | | | | 21.01 | 4 | 9 | 4 | 104.3 | 0 | |
| P27298 | Oligopeptidase A OS=Escherichia coli (strain K12) OX=83333 GN=prlC PE=3 SV=3 | 6 | 1.89E+05 | 6.82E+04 | 7.09E+04 | 1.24E+04 | 1.10E+05 | | | 18.91 | 3 | 5 | 3 | 77.1 | 2 | Oxidation [M190; M259] |
| P0A805 | Ribosome-recycling factor OS=Escherichia coli (strain K12) OX=83333 GN=frr PE=1 SV=1 | 16 | 5.39E+05 | 2.70E+05 | 2.22E+04 | | 1.94E+05 | | | 23.07 | 2 | 8 | 2 | 20.6 | 1 | Oxidation [M83] |
| P0A991 | Fructose-bisphosphate aldolase class 1 OS=Escherichia coli (strain K12) OX=83333 GN=fbaB PE=1 SV=2 | 10 | 5.55E+05 | 1.52E+05 | 1.98E+04 | | 1.80E+04 | | | 4.82 | 2 | 2 | 2 | 38.1 | 1 | Oxidation [M40] |
| P12758 | Uridine phosphorylase OS=Escherichia coli (strain K12) OX=83333 GN=udp PE=1 SV=3 | 15 | 1.84E+06 | 1.14E+06 | 7.51E+05 | 2.53E+05 | | | | 17.84 | 4 | 10 | 4 | 27.1 | 2 | Oxidation [M38; M215] |
| P61889 | Malate dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=mdh PE=1 SV=1 | 11 | 9.28E+05 | 2.79E+05 | 1.32E+05 | 6.37E+04 | | | | 12.49 | 3 | 5 | 3 | 32.3 | 1 | Oxidation [M227] |

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P0AB H7 | Citrate synthase OS=Escherichia coli (strain K12) OX=83333 GN=gltA PE=1 SV=1 | 8 | 6.33 E+0 5 | 4.14 E+0 5 | 2.03 E+0 5 | 6.83 E+0 4 | 9.34 E+0 3 | 2.02 E+0 5 | 1.46 E+0 4 | 7.68 | 3 | 5 | 3 | 48 | 1 | Oxidation [M113] |
| P0A6 U3 | tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG OS=Escherichia coli (strain K12) OX=83333 GN=mnmG PE=1 SV=1 | 4 | | | | | | | | 6.1 | 2 | 2 | 2 | 69 .5 | 2 | Oxidation [M70; M264] |
| P0AE 08 | Alkyl hydroperoxide reductase C OS=Escherichia coli (strain K12) OX=83333 GN=ahpC PE=1 SV=2 | 19 | 9.63 E+0 5 | 7.25 E+0 5 | 1.22 E+0 6 | 1.90 E+0 5 | 1.55 E+0 5 | 2.82 E+0 5 | | 12.1 1 | 3 | 5 | 3 | 20 .7 | 1 | Oxidation [M96] |
| P335 99 | NADH-quinone oxidoreductase subunit C/D OS=Escherichia coli (strain K12) OX=83333 GN=nuoC PE=1 SV=3 | 5 | 4.63 E+0 4 | | | | 5.17 E+0 4 | | | 11.3 4 | 3 | 6 | 3 | 68 .2 | 0 | |
| P698 28 | PTS system galactitol-specific EIIA component OS=Escherichia coli (strain K12) OX=83333 GN=gatA PE=1 SV=1 | 15 | 9.16 E+0 5 | 3.50 E+0 5 | 2.49 E+0 5 | 2.05 E+0 4 | 3.49 E+0 5 | | | 19.4 9 | 2 | 8 | 2 | 16 .9 | 1 | Oxidation [M26] |
| P0A8 V2 | DNA-directed RNA polymerase subunit beta OS=Escherichia coli (strain K12) OX=83333 GN=rpoB PE=1 SV=1 | 3 | 3.73 E+0 5 | 4.14 E+0 4 | 2.97 E+0 4 | | | | | 13.6 5 | 3 | 5 | 3 | 150. 5 | 1 | Oxidation [M704] |
| P0AB P8 | Purine nucleoside phosphorylase DeoD-type OS=Escherichia coli (strain K12) OX=83333 GN=deoD PE=1 SV=2 | 9 | 4.93 E+0 5 | 3.72 E+0 5 | 3.60 E+0 5 | | 1.22 E+0 5 | | | 13.0 2 | 2 | 7 | 2 | 25 .9 | 0 | |
| P371 88 | PTS system galactitol-specific EIIB component OS=Escherichia coli (strain K12) OX=83333 GN=gatB PE=1 SV=2 | 20 | 5.66 E+0 5 | 8.56 E+0 4 | 4.76 E+0 5 | | | 2.38 E+0 4 | 1.57 E+0 4 | 4.38 | 1 | 4 | 1 | 10 .2 | 1 | Oxidation [M48] |
| P0A8 62 | Thiol peroxidase OS=Escherichia coli (strain K12) OX=83333 GN=tpx PE=1 SV=2 | 8 | | 2.98 E+0 4 | | | 6.03 E+0 4 | | | 13.1 3 | 1 | 3 | 1 | 17 .8 | 0 | |
| P066 10 | Thioredoxin/glutathione peroxidase BtuE OS=Escherichia coli (strain K12) OX=83333 GN=btuE PE=1 SV=1 | 10 | | | | | | | | 7.15 | 1 | 2 | 1 | 20 .5 | 0 | |
| P0A6 Z3 | Chaperone protein HtpG OS=Escherichia coli (strain K12) OX=83333 GN=htpG PE=1 SV=1 | 4 | 3.65 E+0 4 | 1.04 E+0 5 | | | 7.94 E+0 4 | | | 4.16 | 2 | 2 | 2 | 71 .4 | 0 | |
| P0A6 Q6 | 3-hydroxyacyl-[acyl-carrier-protein] dehydratase FabZ OS=Escherichia coli (strain K12) OX=83333 GN=fabZ PE=1 SV=1 | 11 | 1.58 E+0 5 | 9.46 E+0 4 | 1.52 E+0 5 | | 2.00 E+0 5 | | | 9.21 | 2 | 5 | 2 | 17 | 0 | |
| P008 64 | Phosphoenolpyruvate carboxylase OS=Escherichia coli (strain K12) OX=83333 GN=ppc PE=1 SV=1 | 2 | 9.77 E+0 5 | 7.20 E+0 4 | | 2.82 E+0 4 | | | | 10.4 9 | 2 | 5 | 2 | 99 | 0 | |
| P321 32 | GTP-binding protein TypA/BipA OS=Escherichia coli (strain K12) OX=83333 GN=typA PE=1 SV=2 | 3 | 8.05 E+0 5 | 1.47 E+0 5 | | | | | | 6.15 | 2 | 3 | 2 | 67 .3 | 0 | |
| P0AG D3 | Superoxide dismutase [Fe] OS=Escherichia coli (strain K12) OX=83333 GN=sodB PE=1 SV=2 | 8 | 9.50 E+0 5 | 3.13 E+0 5 | 1.19 E+0 5 | | 2.78 E+0 5 | | | 10.6 3 | 2 | 5 | 2 | 21 .3 | 0 | |
| P0AB K5 | Cysteine synthase A OS=Escherichia coli (strain K12) OX=83333 GN=cysK PE=1 SV=2 | 4 | 2.25 E+0 5 | 6.59 E+0 4 | 2.97 E+0 4 | | | | | 5.21 | 1 | 2 | 1 | 34 .5 | 0 | |
| P091 52 | Respiratory nitrate reductase 1 alpha chain OS=Escherichia coli (strain K12) OX=83333 GN=narG PE=1 SV=4 | 1 | 2.39 E+0 5 | | | | | | | 7.66 | 2 | 3 | 2 | 140. 4 | 0 | |
| P0A6 Y8 | Chaperone protein DnaK OS=Escherichia coli (strain K12) OX=83333 GN=dnaK PE=1 SV=2 | 3 | 8.27 E+0 5 | | | | 1.08 E+0 4 | | 1.04 E+0 5 | 4.69 | 2 | 2 | 2 | 69 .1 | 0 | |
| P066 12 | DNA topoisomerase 1 OS=Escherichia coli (strain K12) OX=83333 GN=topA PE=1 SV=2 | 2 | 2.96 E+0 4 | | 1.06 E+0 5 | | | | | 2.82 | 1 | 1 | 1 | 97 .3 | 0 | |
| P009 56 | Isoleucine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=ileS PE=1 SV=5 | 2 | 3.33 E+0 5 | 1.51 E+0 5 | 9.24 E+0 4 | | | | | 6.12 | 2 | 5 | 2 | 104. 2 | 0 | |
| P0A8 50 | Trigger factor OS=Escherichia coli (strain K12) OX=83333 GN=tig PE=1 SV=1 | 5 | 3.06 E+0 5 | 1.26 E+0 5 | 1.75 E+0 5 | 1.53 E+0 4 | | | | 4.41 | 2 | 2 | 2 | 48 .2 | 0 | |
| P071 18 | Valine--tRNA ligase OS=Escherichia coli (strain K12) OX=83333 GN=valS PE=1 SV=2 | 1 | 8.50 E+0 4 | 5.48 E+0 4 | 2.37 E+0 5 | | | | | 3.32 | 1 | 1 | 1 | 108. 1 | 0 | |
| P215 99 | Pyruvate kinase II OS=Escherichia coli (strain K12) OX=83333 GN=pykA PE=1 SV=3 | 4 | 1.29 E+0 5 | 6.10 E+0 4 | 2.54 E+0 4 | 1.03 E+0 4 | | | | 4.98 | 1 | 2 | 1 | 51 .3 | 0 | |
| P0AF G3 | 2-oxoglutarate dehydrogenase E1 component OS=Escherichia coli (strain K12) OX=83333 GN=sucA PE=1 SV=1 | 1 | 7.04 E+0 4 | 4.04 E+0 4 | 2.05 E+0 5 | 1.89 E+0 4 | | | | 5.28 | 1 | 2 | 1 | 10 5 | 0 | |
| P0A6 A8 | Acyl carrier protein OS=Escherichia coli (strain K12) OX=83333 GN=acpP PE=1 SV=2 | 12 | 5.33 E+0 5 | 1.88 E+0 5 | 2.13 E+0 5 | | 1.33 E+0 5 | | | 10.3 1 | 1 | 6 | 1 | 8. 6 | 0 | |
| P082 00 | Isocitrate dehydrogenase [NADP] OS=Escherichia coli (strain K12) OX=83333 GN=icd PE=1 SV=1 | 5 | 1.28 E+0 6 | 5.85 E+0 5 | 1.76 E+0 5 | 8.63 E+0 4 | | 4.83 E+0 4 | 4.37 E+0 4 | 7.45 | 2 | 5 | 2 | 45 .7 | 0 | |
| P697 97 | PTS system mannose-specific EIIAB component OS=Escherichia coli (strain K12) OX=83333 GN=manX PE=1 SV=2 | 4 | 1.00 E+0 5 | 6.04 E+0 4 | | 2.38 E+0 4 | 2.79 E+0 4 | | 1.08 E+0 4 | 3.4 | 1 | 1 | 1 | 35 | 1 | Oxidation [M241] |
| P0A7 V8 | 30S ribosomal protein S4 OS=Escherichia coli (strain K12) OX=83333 GN=rpsD PE=1 SV=2 | 4 | 8.30 E+0 4 | 8.58 E+0 4 | | | 9.23 E+0 4 | | | 2.77 | 1 | 1 | 1 | 23 .5 | 0 | |
| P0AE S4 | DNA gyrase subunit A OS=Escherichia coli (strain K12) OX=83333 GN=gyrA PE=1 SV=1 | 2 | 5.87 E+0 4 | | 2.16 E+0 5 | 1.64 E+0 4 | | | | 6.28 | 1 | 3 | 1 | 96 .9 | 0 | |
| P023 58 | 30S ribosomal protein S6 OS=Escherichia coli (strain K12) OX=83333 GN=rpsF PE=1 SV=1 | 8 | 9.01 E+0 4 | 2.85 E+0 5 | 1.32 E+0 4 | | 2.94 E+0 4 | | | 0 | 1 | 1 | 1 | 15 .7 | 0 | |
| P171 69 | Glutamine--fructose-6-phosphate aminotransferase [isomerizing] OS=Escherichia coli (strain K12) OX=83333 GN=glmS PE=1 SV=4 | 2 | | | | | 6.49 E+0 4 | | | 2.61 | 1 | 1 | 1 | 66 .9 | 0 | |
| P0A9 Y6 | Cold shock-like protein CspC OS=Escherichia coli (strain K12) OX=83333 GN=cspC PE=1 SV=2 | 17 | 5.70 E+0 5 | 2.05 E+0 5 | 4.00 E+0 5 | 4.20 E+0 4 | 1.37 E+0 4 | | | 14.2 2 | 1 | 6 | 1 | 7. 4 | 0 | |
| P0A9 C0 | Anaerobic glycerol-3-phosphate dehydrogenase subunit A OS=Escherichia coli (strain K12) OX=83333 GN=glpA PE=1 SV=1 | 3 | | | | | | | | 2.08 | 1 | 1 | 1 | 58 .9 | 0 | |
| P697 83 | PTS system glucose-specific EIIA component OS=Escherichia coli (strain K12) OX=83333 GN=crr PE=1 SV=2 | 9 | 9.04 E+0 4 | | | | 2.78 E+0 4 | | | 2.11 | 1 | 1 | 1 | 18 .2 | 1 | Oxidation [M144] |
| P765 58 | NADP-dependent malic enzyme OS=Escherichia coli (strain K12) OX=83333 GN=maeB PE=1 SV=1 | 1 | 1.88 E+0 5 | 4.63 E+0 4 | 2.78 E+0 4 | 3.95 E+0 4 | | | | 9.2 | 1 | 3 | 1 | 82 .4 | 0 | |
| P0A6 T1 | Glucose-6-phosphate isomerase OS=Escherichia coli (strain K12) OX=83333 GN=pgi PE=1 SV=1 | 2 | 1.37 E+0 5 | 3.41 E+0 4 | 7.41 E+0 4 | | | | | 0 | 1 | 1 | 1 | 61 .5 | 0 | |
| P0A7 K2 | 50S ribosomal protein L7/L12 OS=Escherichia coli (strain K12) OX=83333 GN=rplL PE=1 SV=2 | 9 | 4.79 E+0 4 | 4.97 E+0 4 | 1.63 E+0 4 | | 6.05 E+0 4 | | | 4.72 | 1 | 2 | 1 | 12 .3 | 0 | |
| P003 50 | 6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli (strain K12) OX=83333 GN=gnd PE=1 SV=2 | 3 | 1.77 E+0 5 | | 8.51 E+0 4 | 4.06 E+0 4 | | 5.75 E+0 4 | | 2.15 | 1 | 1 | 1 | 51 .4 | 0 | |
| P0A8 T7 | DNA-directed RNA polymerase subunit beta' OS=Escherichia coli (strain K12) OX=83333 GN=rpoC PE=1 SV=1 | 1 | 1.89 E+0 5 | 7.50 E+0 4 | | | | | | 2.74 | 1 | 1 | 1 | 155. 1 | 1 | Oxidation [M180] |
| P694 41 | Adenylate kinase OS=Escherichia coli (strain K12) OX=83333 GN=adk PE=1 SV=1 | 5 | 4.92 E+0 5 | 2.16 E+0 5 | | | 4.35 E+0 4 | | | 0 | 1 | 1 | 1 | 23 .6 | 0 | |
| P756 78 | Uncharacterized protein YkfA OS=Escherichia coli (strain K12) OX=83333 GN=ykfA PE=4 SV=2 | 3 | 1.87 E+0 7 | 1.30 E+0 8 | 9.37 E+0 6 | 3.55 E+0 5 | 1.80 E+0 6 | | | 0 | 1 | 7 | 1 | 31 .9 | 0 | |
| P0A9 P0 | Dihydrolipoyl dehydrogenase OS=Escherichia coli (strain K12) OX=83333 GN=lpdA PE=1 SV=2 | 4 | 4.32 E+0 6 | 1.24 E+0 8 | 5.07 E+0 6 | 2.37 E+0 5 | 3.46 E+0 4 | 4.47 E+0 4 | 1.54 E+0 5 | 3.42 | 1 | 1 | 1 | 50 .7 | 0 | |
| P0AF G0 | Transcription termination/antitermination protein NusG OS=Escherichia coli (strain K12) OX=83333 GN=nusG PE=1 SV=2 | 8 | | | | 1.40 E+0 5 | 2.00 E+0 4 | | | 2.42 | 1 | 1 | 1 | 20 .5 | 0 | |
| P0AF R4 | Uncharacterized protein YciO OS=Escherichia coli (strain K12) OX=83333 GN=yciO PE=1 SV=1 | 5 | | | | | | | | 1.95 | 1 | 2 | 1 | 23 .2 | 0 | |
| P256 65 | 5-methyltetrahydropteroyltriglutamate--homocysteine methyltransferase OS=Escherichia coli (strain K12) OX=83333 GN=metE PE=1 SV=6 | 1 | 1.70 E+0 5 | 1.04 E+0 5 | 2.16 E+0 5 | 7.18 E+0 4 | 9.93 E+0 3 | 6.05 E+0 4 | 8.00 E+0 4 | 0 | 1 | 2 | 1 | 84 .6 | 0 | |
| P625 17 | Glucans biosynthesis glucosyltransferase H OS=Escherichia coli (strain K12) OX=83333 GN=mdoH PE=1 SV=1 | 2 | 7.87 E+0 5 | 9.53 E+0 5 | 1.94 E+0 6 | 4.63 E+0 5 | 2.94 E+0 4 | 3.69 E+0 5 | 4.44 E+0 5 | 1.7 | 1 | 3 | 1 | 96 .9 | 0 | |
| P242 30 | ATP-dependent DNA helicase RecG OS=Escherichia coli (strain K12) OX=83333 GN=recG PE=1 SV=1 | 2 | | | 5.24 E+0 4 | 2.71 E+0 4 | | 7.69 E+0 4 | 1.37 E+0 5 | 0 | 1 | 2 | 1 | 76 .4 | 0 | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P0AC69 | Glutaredoxin 4 OS=Escherichia coli (strain K12) OX=83333 GN=grxD PE=1 SV=1 | 11 | 8.29E+04 | 4.12E+04 | 7.20E+04 | | | | | 1.98 | 1 | 2 | 1 | 12.9 | 0 |
| P24216 | Flagellar hook-associated protein 2 OS=Escherichia coli (strain K12) OX=83333 GN=fliD PE=1 SV=5 | 6 | | 7.03E+06 | 6.36E+04 | | | | | 5.05 | 1 | 5 | 1 | 48.4 | 0 |
| P0A953 | 3-oxoacyl-[acyl-carrier-protein] synthase 1 OS=Escherichia coli (strain K12) OX=83333 GN=fabB PE=1 SV=1 | 2 | 2.02E+05 | 1.25E+05 | 7.11E+04 | 1.60E+04 | | 3.87E+04 | 1.89E+04 | 2.03 | 1 | 1 | 1 | 42.6 | 0 |
| P33650 | Fe(2+) transporter FeoB OS=Escherichia coli (strain K12) OX=83333 GN=feoB PE=1 SV=1 | 1 | 3.38E+07 | 4.86E+07 | 5.14E+06 | 3.38E+05 | 2.19E+06 | 1.94E+04 | 1.40E+05 | 14.72 | 1 | 10 | 1 | 84.4 | 0 |
| P0A9I3 | Glycine cleavage system transcriptional repressor OS=Escherichia coli (strain K12) OX=83333 GN=gcvR PE=1 SV=2 | 6 | 7.18E+06 | 2.21E+07 | 6.47E+06 | 4.39E+05 | 7.06E+05 | 4.16E+04 | 1.98E+05 | 11.95 | 1 | 5 | 1 | 20.8 | 0 |
| P08201 | Nitrite reductase (NADH) large subunit OS=Escherichia coli (strain K12) OX=83333 GN=nirB PE=3 SV=4 | 1 | 3.42E+05 | 9.41E+04 | 4.34E+04 | 6.14E+04 | | | | 2.03 | 1 | 1 | 1 | 93.1 | 0 |
| P0A6F3 | Glycerol kinase OS=Escherichia coli (strain K12) OX=83333 GN=glpK PE=1 SV=2 | 3 | 2.80E+06 | 1.65E+07 | 6.26E+05 | 3.75E+04 | 2.05E+05 | | 1.39E+04 | 2.5 | 1 | 1 | 1 | 56.2 | 0 |

## ✹ Cnx1 SDS Gel aggregation & expression parameters

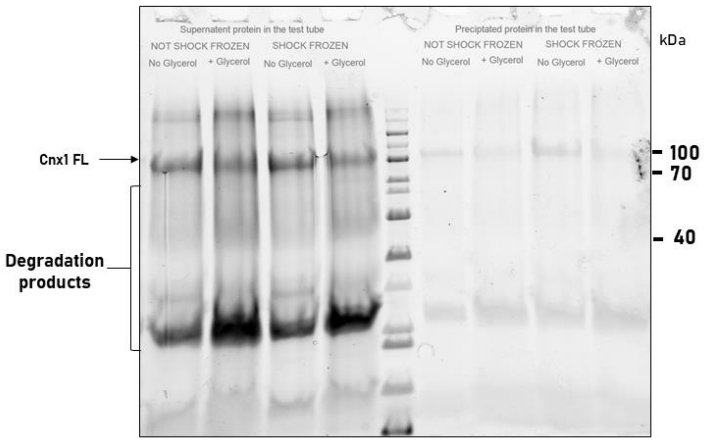



**Figure 70. SDS-PAGE of testing Cnx1 full-length freezing stability.** All Cnx1 full-length proteins were identical, testing the different conditions, including freeze shocking and glycerol addition. Additionally, after centrifugation of proteins, the proteins were transferred to a new eppi, 5x SDS loading buffer was added to the used eppi to evaluate the presence of precipitated proteins. The result showed that with 1 mg/ml protein concentration of Cnx1, all proteins precipitate, regardless of condition.

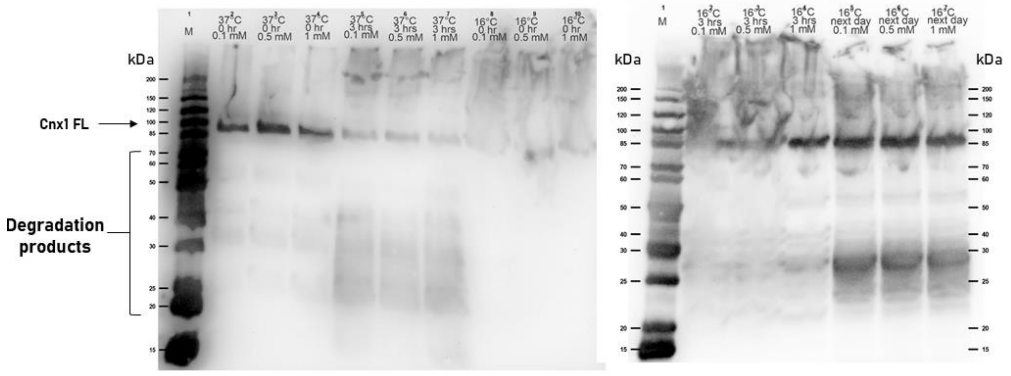



**Figure 71. Anti-strep western blot of the various expression parameters tested for full-length Cnx1.** The western blot was prepared by loading lysate from different conditions. Three different IPTG concentrations were added to each condition,

including 0.1 mM, 0.5 mM, and 1 mM. Samples were taken at 0 hr, 3 hrs and the next day after induction. The temperature either stayed at 37 ◦C or dropped to 16 ◦C. The results show minimal protein expression in general after 3 hrs induction. At 0 hrs of induction, at 37 or 16 ◦C, plenty of proteins are expressed as well as high molecular weight complexes with minimal degradation regardless of IPTG concentration. At 16 ◦C in the next day, degradation products are observed with high molecular weight bands above 200 kDa.

# ✹ Example of TBL file prepared

**Table 26. Example of TBL files prepared for HADDOCK modelling.**

| |
|---|
| assi (segid A and resid 95 and name CA) (segid B and resid 364 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 365 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 378 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 380 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 394 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 401 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 60 and name CA) (segid B and resid 364 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 60 and name CA) (segid B and resid 365 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 60 and name CA) (segid B and resid 394 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 60 and name CA) (segid B and resid 380 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 145 and name CA) (segid B and resid 278 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 40 and name CA) (segid B and resid 31 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 40 and name CA) (segid B and resid 285 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 160 and name CA) (segid B and resid 307 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 45 and name CA) (segid B and resid 307 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 37 and name CA) (segid B and resid 145 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 114 and name CA) (segid B and resid 145 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 442 and name CA) (segid B and resid 364 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 95 and name CA) (segid B and resid 31 and name CA) 04.0 30.0 0.0 |
| assi (segid A and resid 60 and name CA) (segid B and resid 31 and name CA) 04.0 30.0 0.0 |

## ✹ Active and passive residues identified by Cport

Table 27. Active and passive residues for Cnx1E and Cnx1G that is identified by Cport for HADDOCK modelling.

| |
|---|
| **1. For the modified PDB 1UUY or Cnx1G:** |
| **✹ Predicted residues (active residues in HADDOCK):** |
| 14, 15, 16, 19, 20, 21, 23, 24, 25, 28, 29, 79, 82, 86, 87, |
| 88, 111, 112, 115, 116, 117, 118, 122, 138, 139, 140, 141, 142, 143, 145, |
| 146, 149, 443 |
| **✹ Surrounding residues (passive residues in HADDOCK):** |
| 17, 18, 22, 31, 32, 33, 49, 50, 51, 53, 55, 56, 80, 84, 85, |
| 91, 110, 113, 114, 119, 152, 153, 430, 438, 439, 440, 441, 442, 444, 445, |
| 470, 474, 481, 485, 489 |
| **2. For the modified PDB 5G2R or Cnx1E:** |
| **✹ Predicted residues (active residues in HADDOCK):** |
| 15, 16, 17, 18, 19, 21, 23, 37, 65, 66, 69, 78, 83, 87, 95, |
| 104, 107, 108, 120, 122, 123, 148, 149, 150, 233, 266, 269, 270, 292, 450, |
| 451, 454, 456, 458, 471, 513, 518, 530, 555, 557, 558, 668, 701, 702, 704, |
| 705, 709, 710, 712, 732, 733, 759, 760, 833, 834, 835 |
| **✹ Surrounding residues (passive residues in HADDOCK):** |
| 20, 24, 25, 27, 28, 35, 36, 39, 42, 50, 58, 60, 62, 75, 76, |
| 77, 79, 81, 82, 84, 85, 89, 90, 91, 96, 101, 111, 112, 113, 115, |
| 116, 121, 124, 125, 126, 135, 140, 141, 144, 146, 147, 151, 152, 153, 154, |
| 184, 186, 190, 192, 198, 231, 232, 234, 235, 242, 257, 267, 271, 272, 273, |
| 274, 277, 287, 289, 297, 298, 324, 325, 343, 351, 354, 368, 369, 370, 452, |
| 453, 455, 459, 460, 462, 463, 469, 470, 472, 484, 501, 504, 510, 511, 512, |
| 514, 516, 517, 519, 520, 531, 536, 539, 542, 547, 548, 556, 559, 560, 570, |
| 575, 576, 583, 621, 624, 625, 627, 633, 635, 636, 650, 666, 667, 669, 670, |
| 676, 677, 678, 679, 680, 692, 706, 707, 708, 713, 716, 717, 721, 724, 727, |
| 761, 778, 786, 789, 803, 804, 805, 806, 829, 830, 831, 832, 836, 837 |

## ✹ Cnx1 Mutants

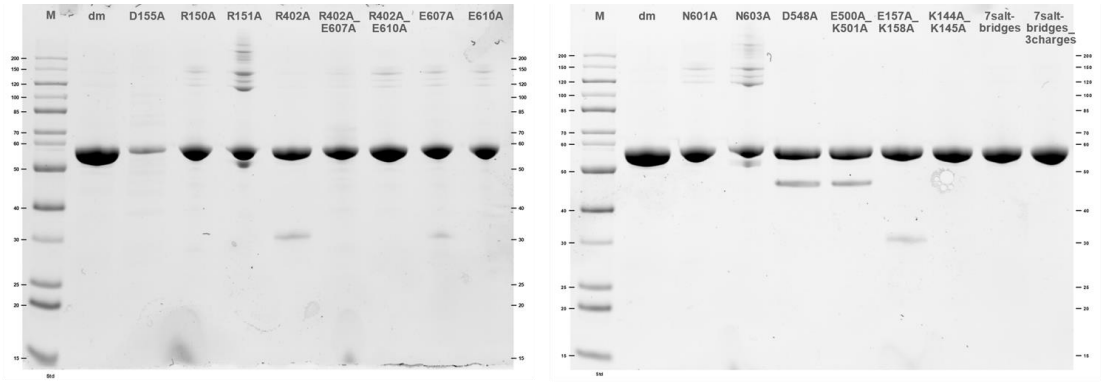



**Figure 72. SDS-PAGE of the Cnx1 mutants.** Protein purity was checked using a 12% gel which was prepared and visualized by TGX stain-free. 10 µg of each protein was loaded (M. Baldauf BA).

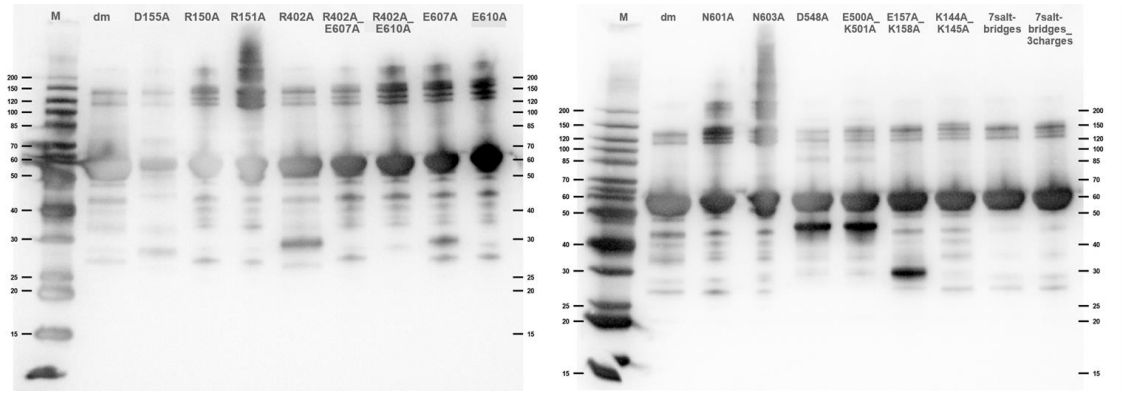



**Figure 73. Anti-strep western blot of the various Cnx1 mutants.** 10 µg of each protein was loaded and blotted using anti-strep tag antibody (M. Baldauf BA).

# ✹ Gel filtration of MoeA-MogA

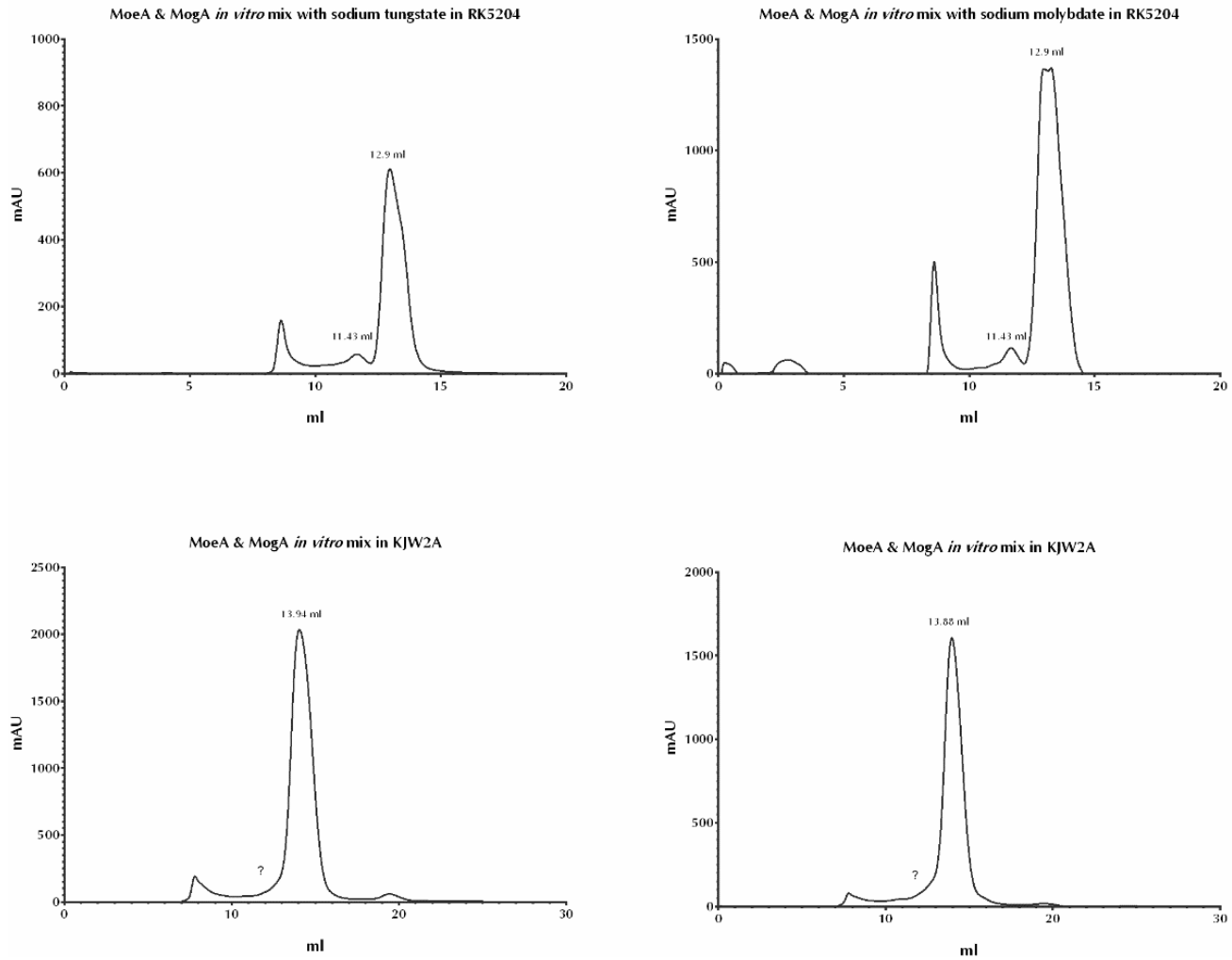



**Figure 74. Gel filtration chromatograms of *in vitro* mixed MoeA-MogA with sodium tungstate, sodium molybdate and in different strains.** The proteins were loaded on a Superdex 200 increase column. All proteins were eluted in 100 mM HEPES, 300 mM NaCl, 1 mM EDTA and 5% glycerol. In the first chromatogram, the MoeA-MogA complex is observed with sodium tungstate added as well as the second chromatogram with sodium molybdate added (expressed in RK5204). For KJW2-A and KJW3-A strains (knock-outs of MoaA and MoaC, respectively), the MoeA-MogA complex is not observed. All main fractions were tested with SDS PAGE.

## ✹ Gephyrin HS gel filtration

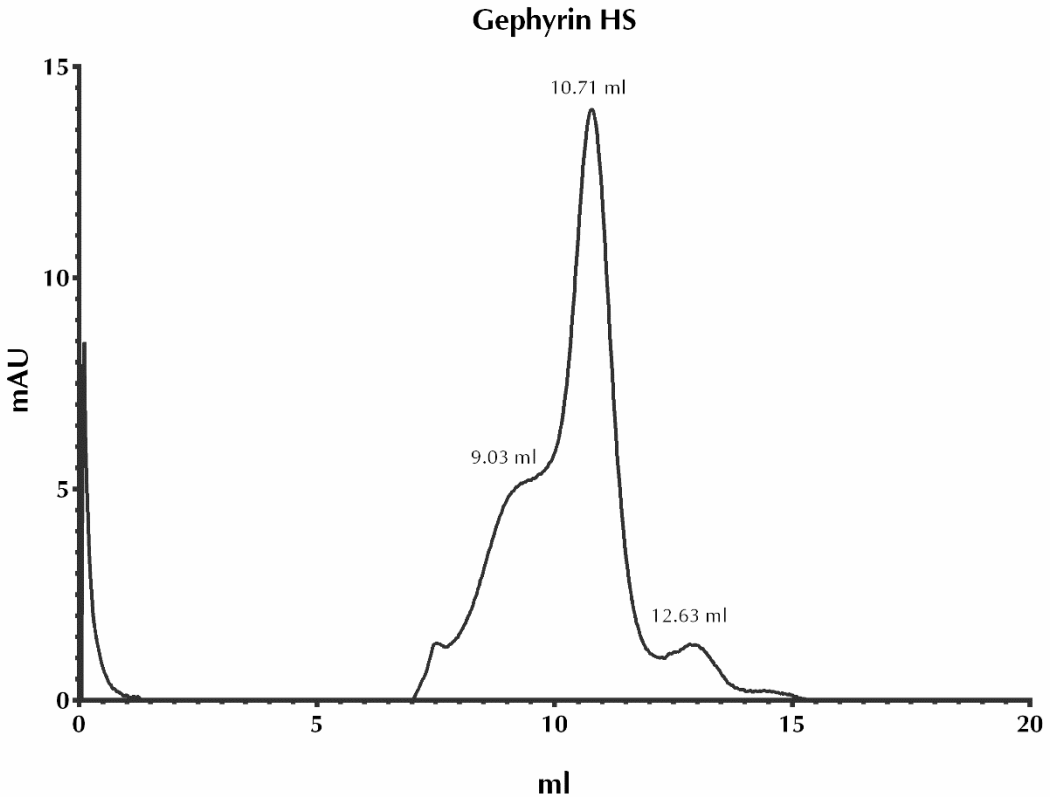



**Figure 75. Gel filtration chromatograms of *Homo sapien* gephyrin isoform 2.** The proteins were loaded on a Superdex 200 increase column. All proteins were eluted in 100 mM HEPES, 300 mM NaCl, 1 mM EDTA and 5% glycerol. The chromatogram shows full-length gephyrin at 12.63 ml, however, at 10.71 ml, a full-length protein complex is present. At 9.03 ml, a higher molecular weight complex is observed. (This chromatogram is generated as part of M. Baldauf bachelor's thesis).

## ✹ Gel filtration of complex forming mutants in comparison to MoeA-MogA

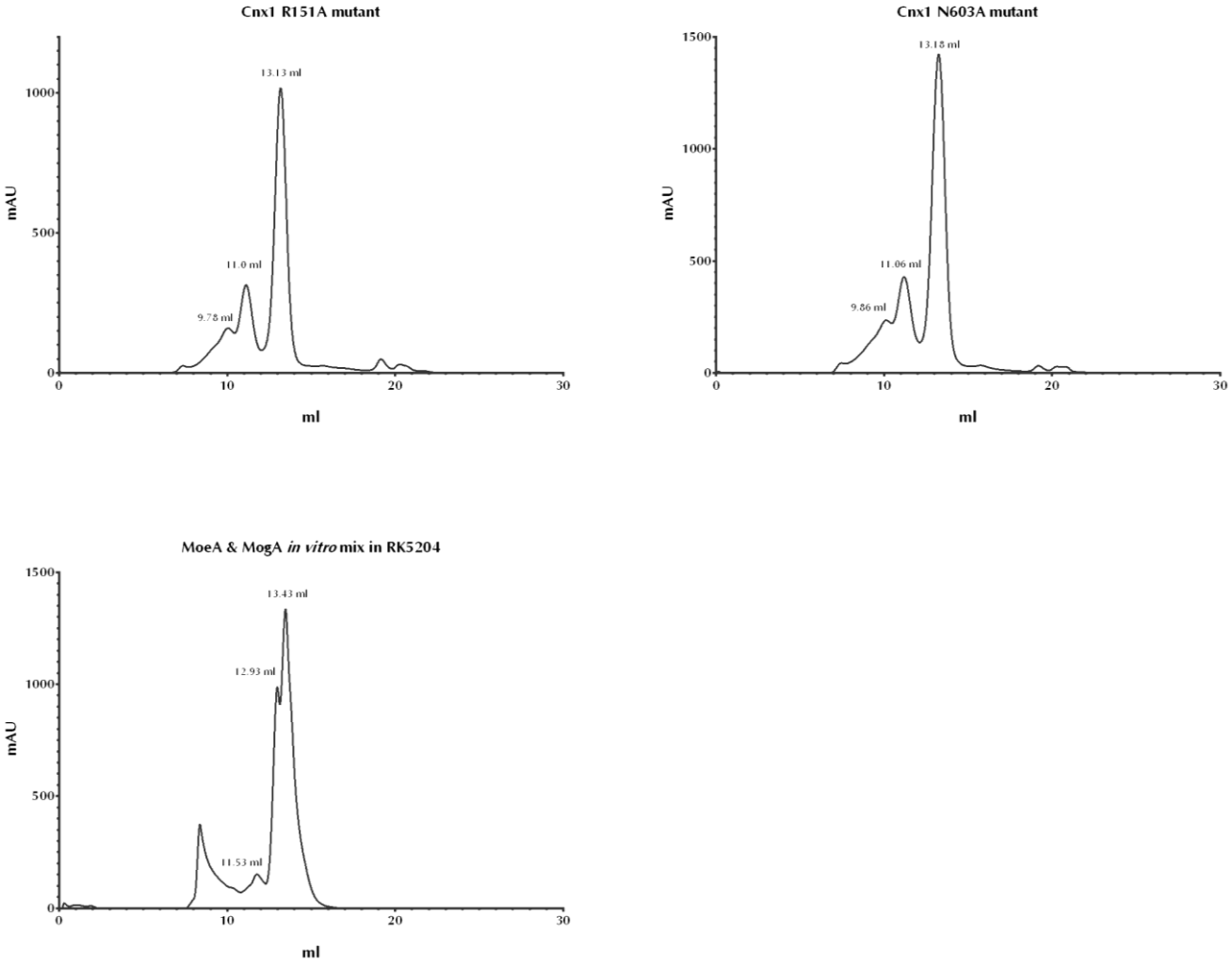



**Figure 76. Gel filtration chromatograms of Cnx1 mutants forming higher molecular weight complexes as well as *in vitro* mixed MoeA-MogA.** The proteins were loaded on a Superdex 200 increase column. All proteins were eluted in 100 mM HEPES, 300 mM NaCl, 1 mM EDTA and 5% glycerol. In the first and second chromatogram, the mutants show complex formation at ≈ 11.0 ml and 9.80 ml. The exact size is correlated to the MoeA-MogA complex showing at 11.53 ml at the last chromatogram. All main fractions were tested with SDS PAGE.

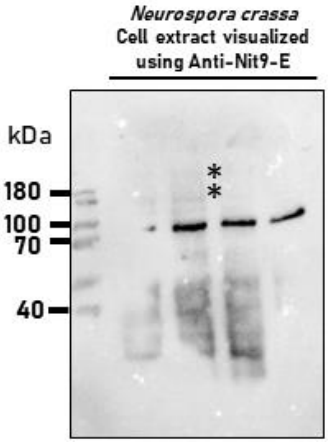



**Figure 77 Western blot of the cell extract of Neurospora crassa using Anti-Nit-9E.** The blot shows the detection of full-length Nit-9 in ≈ 30 µg crude extracts of the mycelia. The primary antibody used is Anti-Nit-9E (Ringel *et al.,* 2013). Full-length protein appears at ≈ 95 kDa. High molecular weight complexes of ≥160 kDa (marked with asterisks).

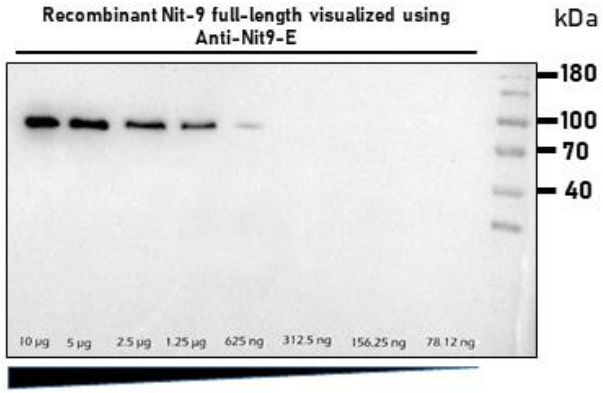



**Figure 78 Western blot of full-length Nit-9 visualized using Anti-Nit-9E**. The blot shows the detection of Nit-9 full-length with different protein amounts loaded (10 µg to 78.12 ng). The used primary antibody is Anti-Nit-9E (Ringel *et al.,* 2013). Full-length protein appears at ≈ 95 kDa.